CHARACTERIZATION OF *E COLI* HFQ STRUCTURE AND ITS RNA BINDING PROPERTIES

A thesis
Presented to
The Academic Faculty

By

Xueguang Sun

In partial Fulfillment
Of the Requirement for the Degree
Doctor of Philosophy in the
School of Biology

Georgia Institute of Technology
May 2006

CHARACTERIZATION OF *E COLI* HFQ STRUCTURE AND ITS RNA BINDING PROPERTIES

Approved by :

Roger M. Wartell, Chair
School of Biology
Georgia Institute of Technology

Stephen C. Harvey
School of Biology
Georgia Institute of Technology

Yury O. Chernoff
School of Biology
Georgia Institute of Technology

Stephen Spiro
School of Biology
Georgia Institute of Technology

Loren D Willimas
School of Chemistry and Biochmestry
Georgia Institute of Technology

Date Approved: November 29 2005

To my family,

for their constant love and support.

ACKNOWLEDGEMENTS

There are many people I would like to thank and acknowledge for their support and help during my five-year Ph.D. study. First and foremost, I would like to thank my advisor, Dr Roger Wartell, for his guidance and assistance throughout this chapter of my career. His constantly open door, scientific insight and perspective, and technical guidance have been integral to furthering my scientific education. He also provided knowledgeable recommendations and multi-faceted support in my personal life and bridged me to a culture which I have never experienced. Without him, it would be impossible to accomplish this thesis work.

I would like to acknowledge Dr. Stephen Harvey, Dr. Yury Chernoff, Dr. Stephen Spiro and Dr. Loren Williams for being on my thesis committee and helpful discussion in structural modeling.

Special thanks go out to Cameron Thomson in IPST, Dr. Andrew Lyon, Dr. Christoph Fahrni in chemistry department and Dr. Xiaodong Cheng in Emory University for readily helping me with using their fluorimeters, and Dr. Nicholas Hud and his student Swapan Jain for volunteering their time of showing me the RNA radio-labeling.

On a personal note, I would like to express my appreciation to the many past and present colleagues in Atlanta and elsewhere, particularly Sixin Yan, Brooke Boudelat-Park, Bin Chen, Hyegjung Park, Hynmin Kim and other fellows in Dr Wartell's lab for getting me through all of my Atlanta ordeals with helping hand and a friendly beverage.

Finally I would like to thank my family for their eagerness to support me in any way they could. I can not begin to express the gratitude I have for my parents and siblings

because of their incredible love and support. With visits, emails, phone calls, and joyous holiday getting together, they have inspired and encouraged me in all of my accomplishments and taught me the value of dedication, hard work, patience and faith. Throughout my life, their words of wisdom reminded me that with effort, anything is possible…even a Ph.D.

.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION

A	adenine
bp	base pair
C	cytosine
CD	circular dichoism
DEPC	diethylpyrocarbonate
dH_2O	deionized water
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide 5'-triphosphate
EDTA	ethylenediaminetetraacetic acid
G	guanine
H_2O_2	hydrogen peroxide
HPLC	high performance liquid chromatography
M	molar
Min	minute
mg	milligram
ml	milliliter
nm	nanometer
nM	nanomolar
OD_{260}	optical density at 260 nm
ODNs	oligodeoxyribonucleotids
PCR	polymerase chain reaction

RBS	ribosomal binding site
RNA	ribonucleic acid
RNase E	ribonuclease E
RNase H	ribonuclease H
T	thymine
TBE	tris-borate EDTA electrophoresis buffer
TE	tris-EDTA buffer
T_m	melting temperature
U	uracil
UV	ultraviolet
°C	degree Celsius
ΔG	Gibbs free energy
ΔH	enthalpy
ΔS	entropy
µg	microgram
µM	micromolar
µl	microliter
ε	extinction coefficient
2’ OMe	2’-O-methyl modified nucleotide

SUMMARY

Hfq is a highly conserved and abundant RNA-binding protein in bacteria. It was originally discovered as a host factor in *Escherichia coli* (also called HF-I) required for bacteriophage Qβ RNA replication. Later it was shown to be a global regulator of bacterial metabolism. Disruption of the *hfq* gene causes a pleiotropic phenotype. The broad impact of the protein appears to stem from its role in regulating the stability and/or translation of mRNAs from a number of genes that respond to environmental stress. Many regulatory non-coding RNAs require Hfq as an essential component in their regulation of mRNA translation. The ability of Hfq to bind to ncRNAs as well as mRNA target sites suggests a cooperative interaction between Hfq, the mRNA site, and the ncRNA. One mRNA affected by several ncRNAs in conjunction with Hfq is the *rpoS* mRNA that encodes the stationary phase sigma factor σ^s of RNA polymerase. Hfq has also been shown to affect the *in vivo* stability and post-transcriptional expression of mRNAs from the *ompA, mutS, miaA,* and *hfq* genes. Hfq is also involved in regulating the addition of poly(A) tails onto mRNAs.

Hfq was recently shown to contain the Sm1 motif, a characteristic of Sm and LSm proteins that function in RNA processing events in archaea and eukaryotes. In this study, comparative structural modeling was used to predict a three-dimensional structure of the Hfq core sequence. The predicted structure aligns with most major features of the *Methanobacterium thermoautotrophicum* LSm protein structure. Conserved residues in Hfq are positioned at the same structural locations responsible for subunit assembly and RNA interaction in Sm proteins. A highly conserved portion of Hfq assumes a structural

fold similar to the Sm2 motif of Sm proteins. The evolution of the Hfq protein was explored by conducting a BLAST search of microbial genomes followed by phylogenetic analysis. Approximately half of the 140 complete or nearly complete genomes examined contain at least one gene coding for Hfq. The presence or absence of Hfq closely followed major bacterial clades. It is absent from high-level clades and present in the ancient Thermotogales-Aquificales clade and all proteobacteria except for those that have undergone major reduction in genome size. Residues at three positions in Hfq form signatures for the beta/gamma proteobacteria, alpha proteobacteria and low GC gram-positive bacteria groups.

A comparative sequence-structure analysis of Hfq genes in bacterial genomes was employed to identify potential RNA binding residues on the surfaces of the Hfq hexamer. A covariance of residue properties at neighboring positions 12 and 39, and conserved surface residues with high propensities at RNA binding sites of RNA-binding proteins suggested potential sites for Hfq-RNA interactions. Gel-mobility shift assay, fluorescence anisotropy, and fluorescence quenching were employed to characterize the binding of wt and eight mutant Hfq proteins with domain II of DsrA ncRNA ($DsrA_{DII}$) and a polyA sequence (A_{18}). Under conditions where Hfq is a hexamer in solution, titrations at high protein and RNA concentrations indicated a 2 Hfq_6: 1 RNA stoichiometry for strong binding complexes of Hfq with $DsrA_{DII}$ or A_{18}. Mutations on the proximal surface of Hfq had a small affect on Hfq binding to A_{18} ($\leq$ 2 fold), while mutations Y25A and K31A on the distal surface decreased affinity to A_{18} by 1000 fold in solution. Mutations F39A and R16A on the proximal surface reduced affinity to $DsrA_{DII}$ by 6-8 fold using the gel-mobility shift assay while other mutations affected the apparent

K_d by ≤ 2 fold. Hfq with a F39A/L12F double mutation partially regained the affinity for $DsrA_{DII}$ lost by the F39A mutation. The latter observation is consistent with the notion that an aromatic residue at either position directly or indirectly assists Hfq-DsrA interaction.

RNA dimerization has been observed to occur for retroviral RNA genomes where two homologous RNA molecules form a complex. This dimerization complex is essential for viral function and is initiated at a stem-loop structure named the dimerization initiation site (DIS). The DIS of HIV-1 is a highly conserved structure with a self-complementary loop sequence that forms a loop-loop 'kissing' complex. We have found a different mechanism of RNA dimer formation occurring for an *in vitro* transcribed 79 nt RNA. This dimer is enhanced in presence of Mg^{2+} and low temperature. RNA structure prediction (using the program Mfold 3.1) suggests that the first 21 nt from the 5' end of the 79 mer is critical for dimerization. A transition of short intramolecular stem-loops to an intermolecular duplex appears to be responsible. Several methods including an RNase H assay, circular dichroism spectroscopy and sedimentation equilibrium centrifugation are consistent with this prediction.

CHAPTER I

INTRODUCTION

Over the past decades, it has become clear that a variety of proteins and small non-coding RNAs have important or essential roles in regulating gene expression at the post-transcriptional level. Many of these regulators are specific in nature (*i.e.* function at a unique target); however, more recently, several post-transcriptional regulators have been characterized that seem to act at many targets and/or have global effects on gene expression. A shining example of such a general post-transcriptional regulator is the bacterial Hfq protein.

Hfq, also called HF-1, is a small thermostable RNA-binding protein. The molecular weight of the monomer is 11,170 atomic mass units (amu), but solution and structural studies indicated it forms a hexamer. It was initially characterized in *Escherichia coli* as a host co-factor required for the replication of Qβ RNA bacteriophage (Franze de Fernandez *et al*., 1968). The *hfq* gene is located at 94.8 min of the *Escherichia coli* chromosome (Kajitani and Ishihama, 1991) and is a part of a complex operon which also includes the *amiB* and *miaA* genes involved in RNA modification as well as the *hflKC* genes, regulators of an ATP-dependent protease, *FtsH*. This superoperon has at least three σ^{32}-dependent 'heat shock' promoters, in addition to four σ^{70}-dependent promoters. The internal 'heat shock' promoters serve to ensure that the cellular level of Hfq is maintained during stress conditions. An intact *hfq* gene has consistently been found to be essential for growth and viability at temperatures greater than 45 ˚C (Tsui *et*

al., 1996). Also, Hfq has been shown to modulate the rate of its own gene expression by causing a decrease in its mRNA stability. This hints at the existence of an autoregulatory circuit in which the accumulation of excess Hfq down-regulates its own synthesis (Tsui *et al.*, 1997).

Cellular distribution and synthesis of Hfq

Analyses of the synthesis rate and the intracellular content of Hfq have revealed that the protein accumulates in proportion to the growth rate in a manner similar to that observed for ribosomes, and Hfq monomers are almost as abundant in cells as ribosomal proteins. Thus, in a rapidly growing bacteria, the cellular content of Hfq is about 50,000-60,000 molecules. During the transition from growth phase to stationary phase, the level of Hfq decreases gradually to about one third of the log phase number (Kajitanni *et al*., 1994; Azam *et al*., 1999). Contrary to these findings, the cellular amount of Hfq has been reported to increases to an appreciable extent in the stationary cell growth phase (Tsui *et al.*, 1997; Vytvytska *et al*., 1998).

Studies carried out to identify the subcellular localization of the protein found that the vast majority of Hfq is present in association with the translational machinery, but a minor fraction appears to be associated with the nucleoid (Kajitani *et al*., 1994; Azam *et al*., 2000). The latter finding might be attributed to a non-specific binding of Hfq to DNA (Takada *et al*., 1997; Azam and Ishihama, 1999) or alternatively, might be explained by the presence of RNA transcripts associated with nucleotide. Hfq has also shown to interact directly with 70S ribosomes through the 30S ribosomal subunit, as well as with

RNA polymerase in an S1-dependent manner (DuBow *et al.*, 1977; Sukhodolets and Garges, 2003).

Pleiotropic roles of Hfq

Disruption of *hfq* gene displays pronounced pleiotropic phenotypes (e.g. decreased growth rate, increased cell length and sensitivity to UV light) and altered patterns of protein synthesis on two-dimensional gels (Tsui *et al*., 1994). Interestingly, some of the phenotypic properties of an *hfq* null mutant match defects caused by muations in *rpoS*, encoding the stationary phase sigma factor σ^s – a global regulator of various stress situations in bacteria (Hengge-Aronis, 2002). Given this, it was soon established that Hfq is essential for efficient translation of the *rpoS* mRNA (Brown and Elliott, 1996; Muffler *et al*., 1996). However, not all the effects of *hfq* knock-out mutants could be attributed to *rpoS* expression, indicating that there are other targets for regulation by Hfq besides *rpoS* (Muffler *et al*., 1997). In line with this, subsequent work revealed that Hfq, directly or indirectly, destabilizes several mRNAs in an RpoS-independent manner. Furthermore, Hfq targets some mRNAs for degradation by binding to their poly (A) tails and stimulating poly (A) adenylation (Hajnsdorf and Regnier, 2000; Le Derout *et al*., 2003). In addition, Hfq associate with a variety of RNA molecules, including the *ompA* mRNA and a family of small non-coding RNAs that function as riboregulators by modulating the stability or translational efficiency of target mRNAs. The importance of Hfq is further underscored by its multiple roles in bacterial physiology, including virulence, bacteriocin production and nitrogen fixation (Vassukueva and Garber, 2002).

The role of Hfq in *RpoS* mRNA translation

Studies on the expression of *rpoS-lacZ* gene fusions in *E.coli* in wild type vs *hfq*$^-$ strains show that Hfq enhances *RpoS* mRNA translation (Cunning *et al.,* 1998). Evidence suggests that Hfq directly or indirectly alters mRNA structure in the region around the start codon. Mutations in the *RpoS* mRNA that reduce the requirement for Hfq *in vivo* imply that this RNA-binding protein is involved in disrupting an RNA secondary structure that sequesters the ribosome binding site (Shine-Dalgarno sequence). *In vivo* studies indicated that Hfq function also requires a site near the 5' end of the *RpoS* mRNA (Brown and Elliott, 1996). The entire leader region prior to the AUG start codon is ~600 nt.

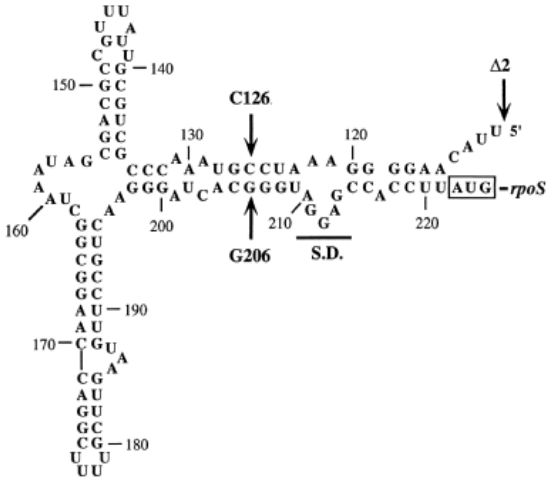


Figure 1.1 Model of secondary structure of 224 nt segment of 5'leader region of *rpoS* mRNA behind translation start point. Structure proposed sequesters RBS (Shine-Dalgarno sequence underlined by S.D.). Mutations C126G and G206C individually enhance translation, decreasing dependence on Hfq. When the complementary mutations restore the base pair, translation is similar to WT (from Brown and Elliott 1996)

A model that does not require additional factors hypothesizes that Hfq binds a segment near the 5' end of the *RpoS* mRNA and simultaneously interacts with a site near

this mRNA's ribosome binding site (Brown and Elliott, 1996). Electron microscopy studies of Hfq binding to Qβ RNA shows looped structures that indicate simultaneous binding of Hfq to two distant sites on this RNA (Schuppli *et al.*, 1997). *In vitro* experiment has demonstrated that Hfq can bind to the *RpoS* mRNA (Lease and Woodson 2004). Regulation of *RpoS* mRNA by Hfq also involve several small regulatory RNAs such as DsrA, OxyS and RprA.

The role of Hfq in small RNA riboregulation

Small regulatory RNAs are a 40-400 nucleotide RNAs encoded by bacterial chromosomes. They do not encode proteins or function as tRNAs or rRNAs. In the past two years, more than 50 non-coding RNAs (ncRNA) in *E. coli* have been identified by systematic computational, microarray and cloning-based screens (Gottesman, 2005). Several of these are known or are proposed to be post-transcriptional regulators. Most of these sRNAs are conserved in related bacteria, indicating that sRNAs are ubiquitous within enterobacteria and arguably in all bacteria (Gottesman, 2002; Wagner and Flarch, 2002).

The importance of Hfq for small RNA riboregulation was first recognized in studies on OxyS RNA, a regulator of the oxidative stress response (Zhang *et al*., 1998). Since then, an increasing number of sRNAs have been shown to associate with Hfq and/or require this protein for post-transcriptional control of target mRNAs (Wassarman, 2002). In a test of 46 known sRNAs found in various searches, 15 were found to bind Hfq tightly; at least 5 other sRNAs were defined by their binding to Hfq (Zhang *et al.*, 2003). The requirement of Hfq for riboregulation may be explained by the stabilization of

some of the sRNAs by Hfq (e.g. DsrA, Spot42 RNA and RyhB) (Sledjeski *et al*., 2001; Masse *et al.,* 2003; Moll *et al*., 2003). Specifically, it has been established that Hfq binding protects the RyhB and DsrA RNAs from cleavage by RNaseE (Mackie, 1998; Masse *et al.,* 2003, Moll *et al*., 2003). This mechanism of action can probably be extended to many of the riboregulators because Hfq binding sites may coincide with recognition sites for RNaseE (Moll *et al.,* 2003). However, Hfq might also function independently of effects on sRNA stability by directly facilitating co-operative RNA-RNA interaction. Insight into this latter function of Hfq has come from studies of the OxyS and Spot 42 RNAs.

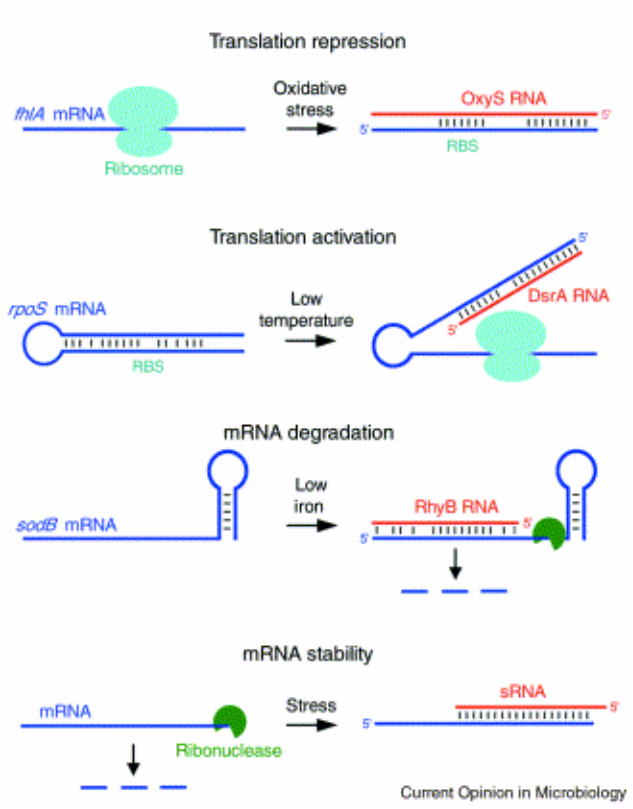


Figure 1.2. Different known and potential regulatory outcomes brought about by sRNA basepairing with mRNAs. sRNAs (red) can repress or activate translation by blocking or promoting ribosome binding to mRNAs (blue). sRNAs also can destabilize or possibly stabilize mRNAs by increasing or decreasing accessibility to ribonucleases. (From Storz 2004)

The OxyS RNA is induced in response to oxidative stress and down-regulates the translation of *rpoS* and *fhlA*. OxyS acts as a translational repressor of *fhlA*. In gel mobility shift assays, Hfq specifically increases the interaction between the OxyS RNA and its target mRNAs. Importantly, the formation of ternary complexes was disrupted by mutations of *fhlA* sequences that are involved in base-pairing with OxyS RNA. Thus, the formation of ternary complexes requires these RNAs to be able to basepair. Hfq does not affect the stability of OxyS RNA but is indispensable for OxyS regulation of *rpoS* and *fhlA in vivo* (Zhang *et al*., 2002).

In contrast to OxyS, DsrA and RprA are two sRNAs which were found independently to upregulate the translation of *rpoS*. Both of them contain a sequence complementary to the upstream stem of the *rpoS* hairpin and activate its translation by basepairing (Lease and Belfort, 2000; Majdalani *et al*., 2002). This is inferred from the fact that mutations in DsrA or RprA that disrupt pairing can be restored to function by compensating mutations in the *rpoS* RNA pairing target. Hfq was found to be essential for this regulation. Comparison of free DsrA to DsrA bound to Hfq by RNase footprinting, circular dichroism spectrum, and thermal melt profiles did not provide evidence that Hfq alters DsrA secondary structure, however, the *in vivo* stability of DsrA was dramatically effected by Hfq (Brescia *et al*., 2003). In wild type strain, DsrA had a half-life of 6-30 min depending on the assay. While in the *hfq*$^{-}$ strain, DsrA was very unstable with a half-life of ~1 min. This sensitivity to degradation in the absence of Hfq is opposite to what was seen with the *ompA, miaA, mutS* and *hfq* mRNAs which are stabilized in an *hfq*$^{-}$ mutant.

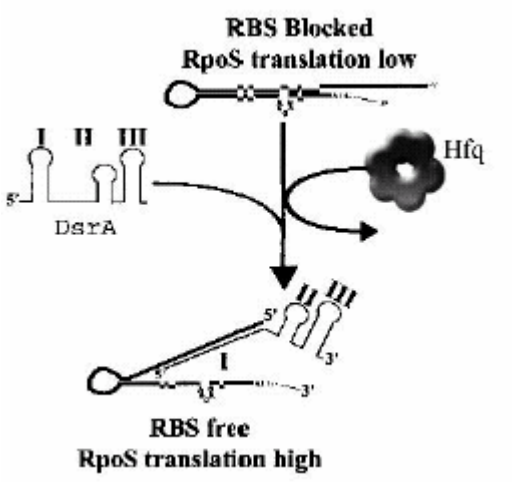


Figure 1.3 Model of Hfq facilitated pairing of domain I of DsrA with region upstream of RBS of *rpoS* mRNA. (from Brescia *et al.,* 2003)

Spot 42 RNA has recently been shown to function by an antisense mechanism to regulate gene expression in the galactose operon (Moller *et al*., 2002b). Spot 42 RNA binds to the translation initiation region of *galK*, the third gene of the operon, to repress translation initiation from this gene. *In vitro*, Hfq facilitate RNA-RNA interaction in a co-operative fashion as demonstrated by the 150-fold increase in complex formation between Spot 42 RNA and its target RNA. In the presence of 1uM Hfq, more than 95% of the 5 nM Spot42 RNA was shifted when adding of 20nM galK RNA in 20mM Na-HEPES pH8, 100mM KCl, 1mM DTT and 1mM Mg^{2+}. Structural probing of free and Hfq-bound Spot 42 RNA suggests that Hfq does not interfere with the structure of Spot 42 RNA. Moreover, Hfq also interacts directly with the *galK* target mRNA region. These findings suggest that Hfq and Spot 42 RNA bind synergistically to the target mRNA, forming a nucleoprotein complex that is held together by multiple RNA-RNA and RNA-protein interactions (Moller *et al*., 2002a; Donahue and Jarrell, 2002).

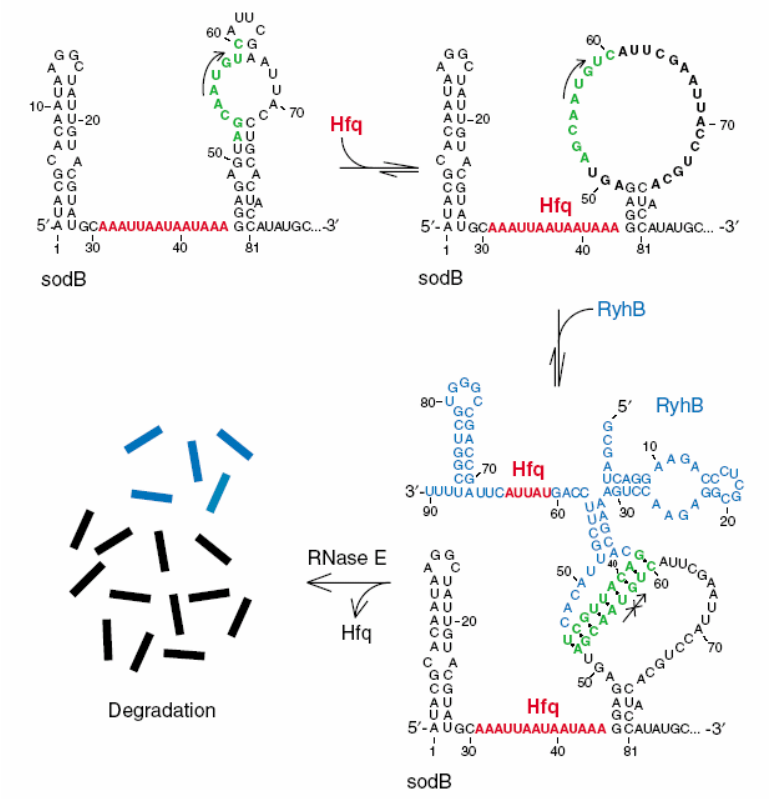


Figure 1.4 Model for *sodB* mRNA−Hfq−RyhB interaction.
(from Geissmann and Touati 2004)

RyhB provides another instructive example of riboregulation in response to particular conditions (Masse and Gottesman, 2002). Fe^{2+} limitation induces the synthesis of this sRNA, Expression of RyhB leads to a decrease in the steady-state level of at least six target mRNAs encoding non-essential iron-binding or iron storage proteins in *E.coli.* Hence, this small regulatory RNA co-operates in optimal utilization of iron when it is scarce. Insight into the mechanism of RyhB action has been acquired from studies of the *sodB* mRNA (Masse *et al.,* 2003a). The work revealed that the target mRNA is rapidly degraded upon induction of RyhB and more surprisingly, that the riboregulator itself becomes highly unstable under conditions in which the target mRNAs are transcribed. Like the Hfq-mediated protection of sRNAs, the degradation of RyhB and the target

RNA is dependent on RNase E. These findings suggest the RyhB acts stoichiometrically rather than catalytically and the RyhB degradation is a consequence of pairing with its target mRNAs.

The role of Hfq in sRNA-independent control

In contrast to the sRNA-dependent control discussed above, it has been suggested that Hfq participates directly in modulating the half-life of the *ompA* mRNA, an exceptionally stable mRNA encoding the major outer membrane protein of *E.coli*. The stability of this transcript correlates inversely with the bacterial growth rate. The half-life of the mRNA is determined by its 5' untranslated region, which contains a stabilizing stem-loop structure as well as recognition sites for RNase E. The efficiency of cleavage at the 5' UTR determines the growth rate-induced changes in *ompA* mRNA turnover. Hfq was identified as a factor present in extracts of slow-growing cells that associates with the 5' UTR of ompA mRNA and is essential for controlling the stability of this RNA (Vytvytska *et al.*, 1998). Thus, in *hfq*$^-$ mutant cells, the half-life of the mRNA is further increased and the growth rate-dependent regulation of its stability is lost. When translation takes place during exponential cell growth (low Hfq), the initiating 30S ribosome subunit binds with the *ompA* mRNA and protects it from RNase E. In stationary phase, Hfq competes with the 30S subunit for the ribosome-binding site of the *ompA* mRNA and prevents formation of the ternary initiation complex (Vytvytska *et al.*, 2000; Moll *et al.,* 2003b). Translation is repressed, RNase E recognition sites are no longer protected by elongating ribosomes, and the mRNA is degraded by the degradosome. A similar mechanism regulating the mRNA life time with Hfq and endonuclease has been

observed in several other cases. For instance, Hfq binding destabilized the *mutS* mRNA, which codes for DNA repair protein recognizing mismatches resulting from methylation (Tsui *et al*., 1997). Hfq binding also destabilizes the *miaA-hfq* cotranscript which codes for a protein modifying ms^2i^6A-37 and i^6A-37 in tRNA and Hfq, respectively, suggesting autoregulation of Hfq expression (Tsui *et al.*, 1994).

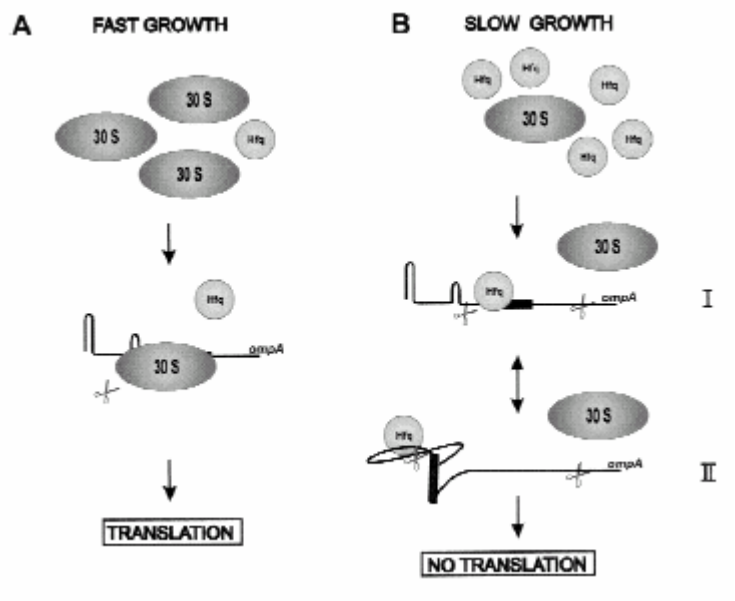


Figure 1.5 Model for growth rate-regulated *ompA* mRNA translation and stability by Hfq. (*A*) Under conditions of fast growth 30S ribosomes are in excess over Hfq, which results in frequent translation initiation of *ompA* mRNA, which, in turn, prevents RNase E cleavages in the 5'-UTR and leads to an enhanced stability of the transcript. (*B*) Under conditions of slow growth, Hfq out competes 30S ribosomes and binds to the *ompA* 5'-UTR. Ribosome binding is either sterically hindered (I) or inhibited by a structural change imposed by Hfq (II). The lack of translation then facilitates RNase E cleavages in the 5'-UTR as well as in the coding region, and results in rapid functional decay of the transcript. Endonucleolytic cleavage is illustrated by scissors. (From Vytvytska *et al.,* 2000)

The role of Hfq in poly(A) metabolism

Prokaryotic mRNA polyadenylation differs strikingly from that in eukaryotic systems. For instance, only 2% of the total mRNA is polyadenylated in *E.coli.* (Cao and

Sarkar, 1992). The size of the poly(A) tail of the bacterial mRNAs depends on competition between poly(A) polymerase and exonucleases, which attack the 3' mRNA end. As a result, poly(A) is commonly 10-40 nt in size (O'Hara *et al*., 1995). Hfq is a cofactor of poly(A) metabolism *in vivo* that affects the fraction of polyadenylated molecules, the distribution of the adenylation sites and the length of poly(A) tails. The higher fraction of polyadenylated molecules found in the presence of Hfq suggests this protein facilitates the recognition of 3' ends by PAPI. A hypothesis was proposed that Hfq is a prokaryotic poly(A)-binding protein, which protects mRNA from ribonucleases (Hajnsdorf and Regnier., 2000). Recently, Mohanty *et al.* showed that inactivation of Hfq leads to the reduction in the ability of PAP I to add poly(A) tails the 3' termini of mRNAs even though PAPI protein levels remain unchanged. Those poly(A) tails that are synthesized in the absence of Hfq are shorter in length. The biosynthetic activity of PNPase in the hfq^- mutant is enhanced and it becomes the primary polynucleiotide polymerase adding heteropolymeric tails almost exclusively to 3' truncated mRNAs. The co-purification of PNPase and Hfq with PAP-I under native conditions indicates a potential complex among these proteins (Mohanty *et al*., 2004). By analogy with eukaryotic systems, the poly(A) tail of RNA was assumed to provide for the interaction of bound translational activities or suppressors (including Hfq) with 5' operators.

The role of Hfq in other bacteria

Hfq is conserved in a wide range of bacteria. The *hfq* homolog encoded by *nrf* gene of *Azorhizobium caulinodans* acts as positive regulator and controls the cell response to two exogeneous signals, environmental nitrogen and oxygen (Kaminski and

Elmerich, 1998). This protein (termed NfrA according to the original nomenclature) was assumed to control the stability and the extent of translation of the *nifA* mRNA, which codes for a transcriptional activator of a gene cascade modulating cell metabolism. The transcript may be stabilized owing to its binding with the Hfq homolog, a change in the transcript secondary structure, and more efficient translation initiation. In the pathogenic bacterium *Yersinia enterocolitica*, the Hfq homolog encoded by *yrp* (*Yersinia* regulator for pleiotropic phenotype) positively regulates the production of thermostable entero-toxin Y-ST, which causes diarrhea in rabbits (Nakao *et al.,* 1995). The regulatory mechanism is still unclear and possibly involves the σ^s subunit of the RNA polymerase. In *Brucella abortus*, Hfq also plays an important role in pathogeneicity (Roberson and Roop 1999). Presumably, its product is responsible for various kinds of stress resistance in the stationary growth phase, including resistance to hydrogen peroxide or low pH. Inactivation of *hfq* changes the *B. abortus* virulence: bacterial cells lose the ability to reproduce in cultured mouse macrophages and are rapidly eliminated from the liver and spleen of artificially infected mice. In addition, a recent paper reported that Hfq together with multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae* (Lenz *et al*., 2004). Hfq mediates interaction between small RNAs and their target mRNA encoding the quorum-sensing master regulators LuxR (*Vibrio harveyi*) and HapR (*Vibrio cholerae*). These interactions typically alter the stability of the target transcripts. Although functional studies of Hfq in bacteria other than *E.coli* are sparse, the highly conserved nature of this protein implies that it is widely used in the regulation of mRNA expression in environmental response pathways. Understanding the molecular

mechanism of Hfq in regulating mRNA expression and its contribution to the survival of pathogenic bacteria may provide important information on host-pathogen interactions.

The RNA binding mechanism of Hfq

Initial characterization of Hfq in the 1970's provided a number of important insights on Hfq-RNA interactions. Results obtained by Franz De Fernandez *et al.* indicated that the protein is predominantly a hexamer in a 0.2 M NH_4SO_4 solution and has a binding preference for poly(rA) and poly(rU) and little binding to poly(rG) and poly(rC) (Franze de Fernandez *et al*., 1972). Other groups demonstrated that Hfq binding to RNA sites can be highly selective. A single Hfq hexamer binds to one site on the ~3800 nt R17 bacteriophage RNA and two sites within the ~4200 nt Qβ RNA to produce maximal replication stimulation. RNAse T1 digestion of these two RNAs with Hfq bound gave three different protected A/U rich fragments 15-26 nt long (Senear and Steitz., 1976)

Studies by deHaseth and Uhlenbeck in 1980 on Hfq binding to $p(A)_n$ oligomers showed that the smallest number of contiguous nucleotides which allowed formation of all favorable Hfq-$(pA)_n$ contacts was 16. The binding affinities of Hfq to $(pA)_n$ oligomers ($12<n<27$) were relatively insensitive to salt concentration from 0.1M to 1.0M Na^+. The data implied only ~2 inonic interactions for $(pA)_{16}$ and suggested that non-ionic interactions dominates polyA-Hfq binding. An experimental result pertinent to a model discussed below compared Hfq binding to linear and circular $(pA)_n$. The equilibrium binding constant of Hfq to circular $(pA)_{18}$ was 9.7 times higher than Hfq binding to linear $(pA)_{18}$, while circular $(pA)_{15}$ bound less well (0.38) than linear $(pA)_{15}$. This suggested to

the authors that Hfq's RNA binding site has a circular geometry 16-18 nt long (deHaseth and Uhlenbeck, 1980a).

Another interesting observation by deHaseth and Uhlenbeck was that a considerable fraction of Hfq formed aggregates larger than a hexamer when the sodium ion concentration decreased below 0.3 M NaCl. Although this did not appear to greatly influence $(pA)_n$ binding, changes in ionic strength between 0.3 M and 0.1 M Na^+ had a major affect on Hfq affinity to Qβ RNA. Hfq creates an RNA loop when it binds to two well separated sites on Qβ RNA and its affinity is much greater than its affinity to short segments containing the individual sites. The results implied that the aggregation status of Hfq plays an important role in its strong affinity to distal sites on a long RNA. The nature of the state of aggregation beyond the hexamer was not eluciated (deHaseth and Uhlenbeck, 1980b).

During the past two years a number of studies have revealed new information on Hfq interactions with RNA, especially in small RNA riboregulation. *In vitro*, Hfq stimulates the pairing of small RNAs with their target messages. Two possible roles for Hfq in this stimulation of pairing have been suggested; both may be true (Storz *et al*., 2004). In one model, interactions between the RNAs and Hfq increase local concentration, aiding RNA:RNA interaction. For instance, one Hfq hexamer may bind to the sRNA and target mRNA simultaneously. Alternatively, one Hfq hexamer may bind the sRNA and a second Hfq hexamer may bind the mRNA. The two Hfq hexamers could be brought together via interactions between the hydrophobic backs of the two hexamers. The second model suggests more of a chaperone role. Hfq binding to a sRNA or to a target may change and/or stabilize RNA structure in such a way that complementary

sequences are more available for pairing. This has been studied *in vitro* in a few instances, Although Hfq binding did not change the structure of RyhB or DsrA, it did change the structure of the RyhB target *sodB* (figure1.4), improving its ability to pair with RyhB. Subtle changes in OxyS structure were detected on Hfq binding. Possibly Hfq also recruits other activities that have more direct roles in changing RNA structure. A recent paper suggests that Hfq itself has ATPase activity and that it associates with ribosomal protein S1 and, through that association, with RNA polymerase; this would suggest significantly more complex roles for Hfq (Sukhodolets and Garges, 2003).

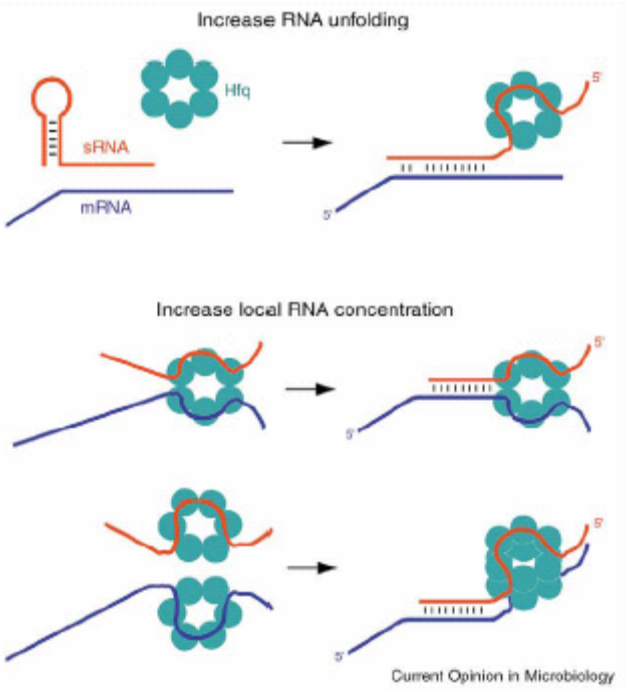


Figure 1.6 Mechanisms by which Hfq might facilitate sRNA-mRNA basepairing. Hfq (aqua ring) may promote RNA unfolding or may increase the local concentrations of the sRNA (red) and its mRNA target (blue).

Objectives

Although Hfq has been studied widely in distinct biological processes, the available structural information is still limited and no crystallographic results were available at the point our work began. Comparative amino acid sequence analysis showed that Hfq is evolutionarily related to the large class of Sm and Sm-like (Lsm) proteins found in eukaryotes and archea. The Sm class of proteins form ring-like heteroheptamers in eukaryotes and are main components of spliceosomal small nuclear ribonucleoproteins (Kambach *et al.,* 1999). Lsm proteins in Archaea form ring-like homoheptamers and appear to also be involved in RNA processing events (Achsel *et al*., 2001). All Sm and Lsm proteins contain a conserved Sm motif, which consists of two segments, Sm1 and Sm2, and an intermediate sequence varying in size and amino acid composition (Achsel *et al*., 2001). The presence of the Sm1 motif sequence in *E. coli* Hfq support the notion that Hfq evolved from the common ancestor with the Sm/Lsm proteins (Collins *et al.*, 2001), although the absence of the Sm2 motif in Hfq makes the structural relationship between Hfq and the Sm protein uncertain. To examine structure-function relationships in Hfq, I predicted the a three-dimensional structure using two web-based homology modeling programs 3D-PSSM and SWISS-MODEL, combined with the secondary structure prediction and sequence analysis. Comparisons of the model with three crystal structures which came out later showed that the model was in an excellent agreement with the crystal structures. In order to determine the pervasiveness of the Hfq gene and related Sm protein in bacteria and explore its evolution, I conducted a BLAST search against all the bacterial genomes available at the NCBI database and built a phylogenetic

tree of Hfq based on the searching result. The presence and absence of the Hfq protein in particular species closely follows the recently redefined major bacterial clades.

Another goal of this thesis research was to understand the molecular mechanism by which Hfq recognize RNA molecules. I searched for the potential RNA binding sites on the surface of the protein using comparative genomic and structural analysis. I generated wild type *E coli* Hfq as well as several mutant proteins with single or double mutation in the residues involved in the hypothetical RNA binding site *in vitro*. Biochemical and biophysical methods such as gel mobility shift assay, CD spectrum, fluorescence quenching and fluorescence anisotropy were employed to determine the influence of these residues on RNA binding and test for predicted protein-RNA contacts. Their binding constants to specific RNA sequences were measured and the interactions governing RNA recognition was examined.

References

Achsel,T., Stark,H. and Luhrmann,R. (2001) The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. *Proc. Natl Acad. Sci. USA*, **98**, 3685–3689

Azam, T.A., and Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli*. *J Biol Chem* **274**: 33105**–**33113.

Azam, T.A., Hiraga, S., and Ishihama, A. (2000) Two types of localization of the DNA-binding proteins within the Escherichia coli nucleoid. *Genes Cells* **5**: 613 626.

Brescia CC, Mikulecky PJ, Feig AL, Sledjeski DD. (2003) Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. *RNA*. **9**:33-43.

Brown, L., and Elliott, T. (1996) Efficient translation of the RpoS sigma factor in *Salmonella typhimurium* requires host factor I, an RNA-binding protein encoded by the hfq gene . *J Bacteriol* **178**: 3763**–**3770.

Cao GJ, and Sarkar N., (1992) Identification of the gene for an Escherichia coli poly(A) polymerase. *Proc Natl Acad Sci USA*. **89**:10380-4.

Collins,B.M., Harrop,S.J., Kornfeld,G.D., Dawes,I.W., Curmi,P.M. and Mabbutt,B.C. (2001) Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs. *J. Mol. Biol.*, **309**, 915–923

Cunning C, Brown L, Elliott T., (1998) Promoter substitution and deletion analysis of upstream region required for rpoS translational regulation. J Bacteriol. **180**: 4564-70.

Donahue, W.F., and Jarrell, K.A. (2002) A BLAST from the past: ancient origin of human Sm proteins. *Mol Cell* **9**: 7**–**8.

DuBow, M.S., Ryan, T., Young, R.A., and Blumenthal, T. (1977) Host factor for coliphage Q beta RNA replication: presence in procaryotes and association with the 30S ribosomal subunit in Escherichia coli. *Mol General Genet* **153**: 39 43

Franze de Fernandez. M.T., Eoyang, L., and August, J.T. (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature* **219**: 588 590.

Franze de Fernandez MT, Hayward WS, August JT. (1972) Bacterial proteins required for replication of phage Q ribonucleic acid. Pruification and properties of host factor I, a ribonucleic acid-binding protein. *J Biol Chem*. **247**:824-31.

Geissmann TA, Touati D., (2004) Hfq, A new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J*. **23**:396-405.

Gottesman, S. (2002) Stealth regulation: biological circuits with small RNA switches. *Genes Dev* **16**: 2829–2842.

Gottesman S. (2004) The small RNA regulators of Escherichia coli: roles and mechanisms. *Annu Rev Microbiol* **58**:303-28.

Hajnsdorf, E., and Regnier, P. (2000) Host factor Hfq of *Escherichia coli* stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proc Natl Acad Sci USA* **97**: 1501–1505

de Haseth, P.L., and Uhlenbeck, O.C. (1980a) Interaction of *Escherichia coli* host factor protein with Q beta ribonucleic acid. *Biochemistry* **19**: 6146 6151.

de Haseth, P.L., and Uhlenbeck, O.C. (1980b) Interaction of *Escherichia coli* host factor protein with oligoriboadenylates. *Biochemistry* **19**: 6138–6146.

Hengge-Aronis, R. (2002) Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase. *Microbiol Mol Biol Rev* **66**: 373 395.

Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Luhrmann,R., Li,J. and Nagai,K. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, **96**, 375–387.

Kajitani, M., Kato, A., Wada, A., Inokuchi, Y., and Ishihama, A. (1994) Regulation of the Escherichia coli Hfq gene encoding the host factor for phage Q-**Beta**. *J Bacteriol* **176**: 531–534.

Kaminski PA, Elmerich C. (1998) The control of Azorhizobium caulinodans nifA expression by oxygen, ammonia and by the HF-I-like protein, NrfA. *Mol Microbiol*. **28,** 603-13.

Lease RA, Belfort M. (2000) A trans-acting RNA as a control switch in Escherichia coli: DsrA modulates function by forming alternative structures. *Proc Natl Acad Sci USA*. **97**:9919-24.

LeDerout, J., Folichon, M., Briani, F., Deh, G., Rgnier, P., and Hajnsdorf, E. (2003) Hfq affects the length and the frequency of short oligo(A) tails at the 3' end of *Escherichia coli* rpsO mRNAs. *Nucleic Acids Res* **31**: 4017–4023.

Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL. (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae*. Cell*. **118**:69-82.

Mackie, G.A. (1998) Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **395**: 720 723.

Majdalani N, Hernandez D, Gottesman S. (2002) Regulation and mode of action of the secondsmall RNA activator of RpoS translation, RprA. *Mol Microbiol*. **46**: 813-26.

Masse, E., and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA* **99**: 4620 4625.

Masse, E., Escorcia, F.E., and Gottesman, S. (2003a) Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev* **17**: 2374 2383.
Mohanty, B. K. Maples V.F. and S. R. Kushner. (2004). The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*. *Mol.Microbiol*., **54**(4) 905.

Moll, I., Leitsch, D., Steinhauser, T., and Blaci, U. (2003b) RNA chaperone activity of the Sm-like Hfq protein. *EMBO Rep* **4**: 284 289.

Moller, T., Franch, T., Hojrup, P., Keene, D.R., Bachinger, H.P., Brennan, R.G., and Valentin-Hansen, P. (2002a) Hfq: a bacterial Sm-like protein that mediates RNA–RNA interaction. *Mol Cell* **9**: 23–30.

Moller, T., Franch, T., Udesen, C., Gerdes, K., and Valentin-Hansen, P. (2002b) Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. *Genes Dev* **16**: 1696 1706.

Muffler, A., Fischer, D., and HenggeAronis, R. (1996) The RNA-binding protein HF-I, known as a host factor for phage Q beta RNA replication, is essential for Rpos translation in *Escherichia coli*. *Genes Dev* **10**: 1143 1151.

Muffler, A., Traulsen, D.D., Fischer, D., Lange, R., and HenggeAronis, R. (1997) The RNA-binding protein HF-I plays a global regulatory role which is largely, but not exclusively, due to its role in expression of the sigma(s) subunit of RNA polymerase in Escherichia coli. *J Bacteriol* **179**: 297–300.

Nakao H, Watanabe H, Nakayama S, Takeda T. (1995) yst gene expression in Yersinia enterocolitica is positively regulated by a chromosomal region that is highly homologous to Escherichia coli host factor 1 gene (hfq). *Mol Microbiol*. **18** :859-65.

O'Hara EB, Chekanova JA, Ingle CA, Kushner ZR, Peters E, Kushner SR. (1995) Polyadenylylation helps regulate mRNA decay in *Escherichia coli*. *Proc Natl Acad Sci USA*. **92**:1807-11.

Robertson GT, Roop RM Jr. (1999) The Brucella abortus host factor I (HF-I) protein contributes to stress resistance during stationary phase and is a major determinant of virulence in mice. *Mol Microbiol*. **34:** 690-700.

Schuppli D, Miranda G, Tsui HC, Winkler ME, Sogo JM, Weber H. (1997) Altered 3'-terminal RNA structure in phage Qbeta adapted to host factor-less Escherichia coli. *Proc Natl Acad Sci US A*. **94**:10239-42.

Senear, A.W., and Steitz, J.A. (1976) Site–specific interaction of Qbeta host factor and ribosomal protein S1 with Qbeta and R17 bacteriophage RNAs. *J Biol Chem* **251**: 1902 1912.

Sledjeski, D.D., Whitman, C., and Zhang, A. (2001) Hfq is necessary for regulation by the untranslated RNA DsrA. *J Bacteriol* **183**: 1997–2005.

Storz G, Opdyke JA, Zhang A. (2004) Controlling mRNA stability and translation with small, noncoding RNAs. Curr Opin Microbiol. **7**:140-144.

Sukhodolets, M.V., and Garges, S. (2003) Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry* **42**: 8022–8034.

Tsui, H.C.T., Leung, H.C.E., and Winkler, M.E. (1994a) Characterization of broadly pleiotropic phenotypes caused by an Hfq insertion mutation in Escherichia coli K-12. *Mol Microbiol* **13**: 35–49.

Tsui HC, Winkler ME., (1994b) Transcriptional patterns of the mutL-miaA superoperon of Escherichia coli K-12 suggest a model for posttranscriptional regulation. *Biochimie* **76**:1168-77.

Tsui, H.C.T., Feng, G., and Winkler, M.E. (1996) Transcription of the mutL repair, miaA tRNA modification, hfq pleiotropic regulator, and hflA region protease genes of Escherichia coli K-12 from clustered E sigma (**0**)-specific promoters during heat shock. *J Bacteriol* **178**: 5719–5731.

Tsui, H.C.T., Feng, G., and Winkler, M.E. (1997) Negative regulation of mutS and mutH repair gene expression by the Hfq and RpoS global regulators of *Escherichia coli* K-12. *J Bacteriol* **179**: 7476–7487.

Vytvytska, O., Jakobsen, J.S., Balcunaite, G., Andersen, J.S., Baccarini, M., and von Gabain, A. (1998) Host factor I, Hfq, binds to Escherichia coli ompA mRNA in a growth rate dependent fashion and regulates its stability. *Proc Natl Acad Sci USA* **95**: 14118 14123.

Vytvytska, O., Moll, I., Kaberdin, V.R., von Gabain, A., and Blasi, U. (2000) Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. *Genes Dev* **14**: 1109 1118.

Wagner, E.G., and Flardh, K. (2002) Antisense RNAs everywhere *Trends Genet* **18**: 223–226.

Wassarman, K.M. (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* **109**: 141–144.

Zhang, A.X., Altuvia, S., Tiwari, A., Argaman, L., Hengge-Aronis, R., and Storz, G. (1998) The OxyS regulatory RNA represses Rpos translation and binds Hfq (HF-I) protein. *EMBO J* **17**: 6061 6068.

Zhang, A.X., Wassarman, K.M., Ortega, J., Steven, A.C., and Storz, G. (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell* **9**: 11–22.

Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol*. **50:**1111-24.

CHAPTER II

STRUCTURE PREDICTION AND PHYLETIC ANALYSIS OF HFQ

Introduction

Amino acid sequence analysis of Hfq has shown that the N-terminal portion of Hfq is highly conserved among a number of bacteria and shares a strong homology with the Sm1 motif of Sm and Sm-like (LSm) proteins found in eukaryotes and archea (Zhang *et al.,* 2002; Moller *et al*., 2002). These results suggest that Hfq is an ancestral Sm protein. Sm proteins are essential components of the small nuclear ribonucleoproteins (snRNPs) that form spliceosomes (Lerner *et al.,* 1979; Luhrmann *et al*., 1990). Sequence comparisons of Sm proteins from a range of species showed that the Sm motif was comprised of two conserved regions, Sm1 and Sm2, separated by a region varying in length and sequence (Hernann *et al.,* 1995). Biochemical and crystallographic studies (Kambach *et al*., 1999; Walk *et al*., 2001) have demonstrated that the Sm motif dictates a common folding domain that enables Sm proteins to assemble onto a uridine-rich region of snRNAs and form a ring-like heteroheptamer. Formation of this core structure is essential for the stability and function of the snRNPs (Fischer *et al*., 1993).

Searches of eukaryotic genome databases have shown that a large number of proteins contain the Sm sequence motif (Seraphin, 1995; Salgado-Garrido *et al.,* 1999). Some of these proteins are very similar to the originally characterized splicesomal Sm proteins, and others are referred to as Sm-like proteins (LSm). Analysis of archaeal

genomes also revealed the presence of ORFs that encode LSm proteins (Salgado-Garrido *et al.,* 1999). Biochemical and crystal studies of several archeal LSm proteins exhibit properties similar to their counter parts in eukaryotes. They bind to RNA with oligo(U) sequences, and assemble a heptameric ring around the RNA (Achsel *et al.,* 2001; Toro *et al.,* 2001). A comparison of the monomer subunits in the crystal structures of eucaryotic Sm proteins that form dimers with the monomer subunits of archeal LSm proteins that form homodimers and heptamers (Mura *et al.,* 2001; Collins *et al.,* 2001) show strong similarities. Each subunit has a short alpha helix followed by five interwoven beta strands separated by short loops.

The presence of the Sm1 motif sequence in *Escherichia coli* Hfq and the ability of Hfq to form a hexameric ring and bind RNA support the notion that it is evolutionarily related to the Sm family of proteins. However the absence of the Sm2 motif in Hfq makes the structural relationship of Hfq with the known structures of Sm proteins uncertain. Here secondary structure prediction, amino acid solvation properties, and three-dimensional threading algorithms were used to predict a three dimensional structure for the amino terminal domain of Hfq. The predicted structure fits very well with the major features of C_α backbone of *Methanobacterium thermoautotrophicum* Lsm protein. The Sm1 motif sequence in Hfq is structurally aligned with its counterpart in the archeal Sm protein, and a highly conserved portion of the Hfq sequence assumes a structural fold similar to that of the Sm2 motif of archeal and eucaryotic Sm proteins. Highly conserved residues of Hfq are also located in the same structural region that is responsible for subunit assembly in the Sm proteins. This strong structural homology supports the hypothesis that Hfq is an ancestral Sm protein and contributes confidence in its predicted

3-D structure. Our 3D model of the *E.coli* Hfq is in excellent agreement with the crystal structure of the *Staphylococcus aureus* Hfq (Schumacher *et al.*, 2002)

The presence of the Hfq protein in bacteria was explored by BLAST searches against bacterial genomes available in the NCBI databases. Approximately half of the bacterial genomes examined contain an Hfq protein based on strong amino acid sequence similarity, protein sequence length and amino acid conservation pattern. Phyletic distribution of Hfq indicates that it is an ancient protein. We obtained no evidence that Hfq might be a subject of lateral gene transfer and conclude that gene loss played a major role in its evolution. The bacterial species in which Hfq was absent were highly correlated with specific taxonomic or lifestyle trends.

Materials and Methods

Sequence analysis of Hfq

Non-redundant database searches were performed by the computer program PSI-BLAST (Altschul *et al*., 1997) (http://www.ncbi.nlm.nih.gov/blast/psiblast.cgi), using the amino acid sequence of *E coli* Hfq (GI 16131994) as the primary query sequence. The inclusion threshold (E-value) employed was 0.01. A multiple alignment was constructed by the CLUSTAL W program (Thompson *et al.,* 1997) at the EBI server (http://www.ebi.ac.uk/clustalw/) using the output of the PSI-BLAST search.

Secondary structure prediction of Hfq

The secondary structure was predicted using the consensus method Jnet (Cuff *et al.,* 1998) at the EBI server (http://jura.ebi.ac.uk:8888). Jnet is a neural net work prediction algorithm that applies multiple sequence alignments, along with PSI BLAST and HMM profiles. First, an automatic alignment method was used. This alignment method took the target sequence and scanned it against the non-redundant database. Hits were then post filtered with the program SCANPS using length dependent cut-offs. The resulting sequences were aligned with CLUSTAL W, and further modified to ensure there were no gaps in the target sequence. Finally, four secondary structure prediction algorithms were run and a consensus secondary structure generated.

Fold recognition and 3D modeling

The 3D-PSSM (Position Specific Scoring Matrix) program was employed from the web-server (http://www.bmm.icnet.uk/servers/3dpssm/) (Kelley *et al.,* 2000) to search for proteins with structural homology to Hfq. This server provides 3-D structural information about the backbone of a query protein by scoring the relationship between the residues of a query sequence with the residues of a homologous protein of known structure. The query protein is scanned against a library composed of proteins with known crystal structures and scored for compatibility using several scoring components. These include amino acid sequence profiles built from relatively close homologues, more general profiles containing more remote homologues, matching of secondary structure elements, and matching the propensities of residues to occupy varying levels of solvent

accessibility. Known protein structures within the database with significant homology to the query sequence are used to produce closest fit alignments between query sequence and target structures that maximize position specific scores. The top 20 structural alignments to the query sequence are produced, each illustrating regions of similarity and differences.

The SWISS-MODEL comparative protein modeling server (Guex *et al.*, 1997) (http://www.expasy.ch/swissmod/) was employed to generate a 3-D model of Hfq based on the structural alignment of its sequence to the highest scoring template structure determined by 3-D PSSM. In the initial step of modeling the Hfq query sequence was modified in order to accommodate the four and six residue segments absent from Hfq when compared to the best template structure; the *Methanobacterium thermoautothrophicum* LSm_{α} protein (Figure 2.3A and 2.3B). Residues were inserted into the Hfq sequence at segments marked by dashes shown in Figures 2.3C and 2.3D to produce a query sequence that would match the length of the *Methanobacterium thermoautothrophicum* template sequence. This modified Hfq query sequence was submitted to the server and SWISS-MODEL produced a predicted structure. The query sequence was then changed to the correct Hfq sequence by replacing the inserted residueswith gaps and the Hfq sequence and template sequence resubmitted to the server in 'Optimize Mode' after aligning the gaps to the template sequence as indicated by the 3D-PSSM model.

Results and Discussion

Sequence similarity of Hfq and Sm proteins

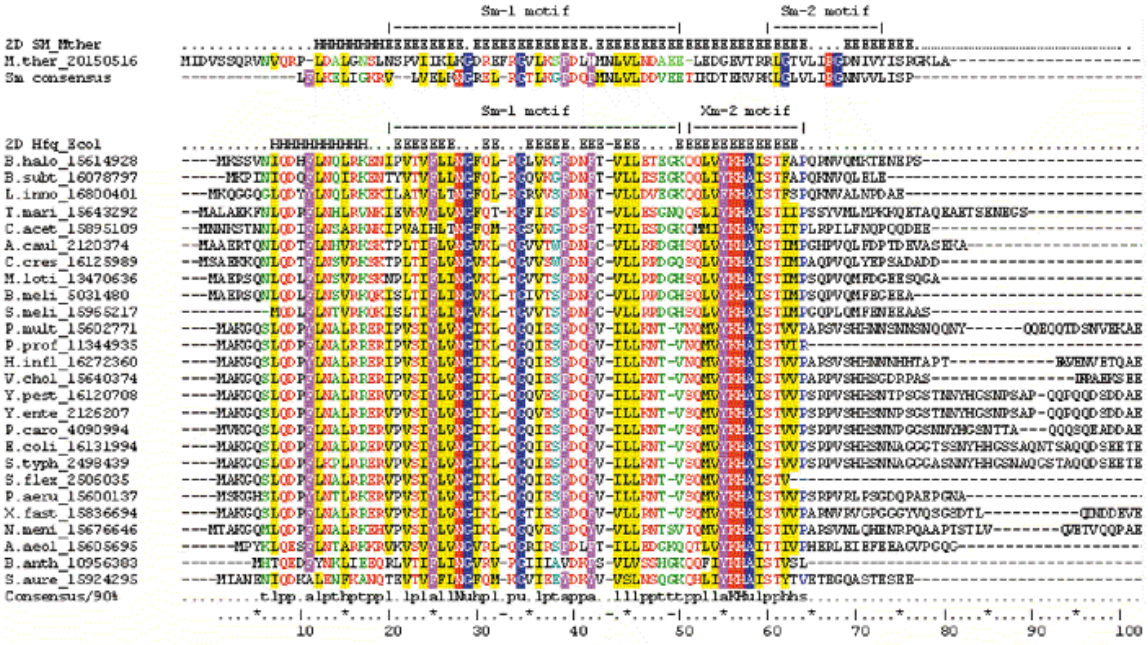


Figure 2.1 Multiple alignment of Hfq proteins from 26 bacterial genomes compared with the LSm protein from *M.thermoautotrophicum* and a consensus sequence for Sm proteins (shown above the alignment). Known secondary structure for the LSm protein and predicted structure for Hfq proteins are shown above the corresponding sequences: H, α helix; E, β strand. The Sm1 and Sm2 motifs of the Sm protein and the Sm1 and Xm2 motifs of Hfq proteins are shown. The 90% consensus shown below the alignment was derived using the following amino acid groupings. Positively charged residues (RKH) are shown as white letters on a red background; polar residues (p, KRHEDQNST) are shown as red letters; turn-like residues (t, ACDEGKNQRST) are green letters; bulky hydrophobic residues (h, ACLIVMHYFW) and the aliphatic subset of these type residues (l, LIVM) have a yellow background; aromatic residues (a, FHWY) are white letters with a purple background; small residues (s, ACDGNPSTV) are blue letters; tiny (u, AGS) are white letters with a blue background. Sequences are denoted by the species abbreviation followed by GI number. Species abbreviations: M.ther, *M.thermoautotrophicum*; B.halo, *Bacillus halodurans*; B.subt, *Bacillus subtilis*; L.inno, *Listeria innocua*; T.mari, *Thermotoga maritima*; C.acet, *Clostridium acetobutylicum*; A.caul, *Azorhizobium caulinodans*; C.cres, *Caulobacter crescentus*; M.loti, *Mesorhizobium loti*; B.meli, *Brucella melitensis biovar Abortus*; S.meli, *Sinorhizobium meliloti*; P.mult, *Pasteurella multocida*; P.prof, *Photobacterium profundum*; H.infl, *Haemophilus influenzae*; V.chol, *Vibrio cholerae*; Y.pest, *Yersinia pestis*; Y.ente, *Yersinia enterocolitica*; P.caro, *Pectobacterium carotovorum*; E.coli, *E.coli*; S.typh, *Salmonella typhimurium*; S.flex, *Shigella flexneri*; P.aeru, *P.aeruginosa*; X.fast, *Xylella fastidiosa*; N.meni, *Neisseria meningitidis*; A.aeol, *Aquifex aeolicus*; B.anth, *Bacillus anthracis*; S.aure, *S.aureus.*

A search of the non-redundant database (NCBI) using PSI-BLAST was carried out with the *Escherichia coli* Hfq amino acid sequence as the query sequence. Twenty-five similar ($P< e^{-5}$) sequences were detected from a range of bacterial species. Multiple alignment of these sequences is shown in Fig. 2.1. Hfq proteins are highly conserved in their N-terminal halves of the molecules. This conserved domain corresponds to residues 7-64 in *E col*i Hfq. In contrast, the C-termini of Hfq proteins vary greatly among the different species. In some instances it is totally absent, e.g., the 57 amino acid Hfq protein of *Bacillus anthracis*. This result implies that the C-terminal region might not play a significant role in the major function(s) of Hfq and attention was focused on the amino terminal region. This hypothesis is supported by a recent study showing that the Hfq homologue of *Pseudomonas aeruginosa*, consisting of 82 amino acids from the N-terminal end, can functionally replace *Escherichia coli* Hfq for phage Qβ replication and for *rpoS* expression (Sonnleitner *et al.*, 2002).

Two conserved motifs are observed in Hfq. The first motif, Sm1, is a counterpart of the Sm1 motif found in archaeal and eukaryotic Sm and LSm proteins (Hermann et al., 1995). From Figure 2.1, it can be seen that the Sm1 sequence is well aligned to residues 20–52 of the *E.coli* Hfq sequence. As noted previously (Moller *et al*., 2002), the Sm2 motif of archaeal and eukaryotic proteins does not appear to have a counterpart in Hfq. However, Hfq does have an additional conserved region, YKHA, following the Sm1 motif. The relationship of the YKHA motif of Hfq with the Sm2 motif of Sm proteins was explored by generating a structural model of Hfq.

Comparison of Predicted Secondary and 3-D Structure of Hfq with Sm Protein

The secondary structure of the consensus Hfq sequence was predicted by JPRED2 (Cuff *et al.,* 1998). Figure 2.1 shows that Hfq is β -sheet-rich structure with an α helix at the N-terminus. All the predicted secondary structure elements fall in the region that is conserved in the multiple alignment. In contrast, no secondary structure elements were predicted in the C-terminus of Hfq. Crystal structures of several Sm proteins show a common fold for the Sm motif. The fold contains an N-terminal helix, followed by five segments of β strands (Kambach *et al.,* 1999; Toro *et al*., 2001; Mura *et al*., 2001; Collins *et al*., 2001). Strands β1, β2 and β3 are part of the Sm1 motif, whereas the Sm2 motif corresponds to β4 and β5 strands (shown in Fig. 2.1). The topology of the secondary structure elements in an Sm protein is schematically shown in Figure 2.2 A and B. Strands β2, β3 and β4 are strongly bent to allow the formation of the hydrophobic core. The structural plasticity needed for such a high degree of curvature in the β1 strand is provided by several strictly conserved glycines that occur near the pivot points (Gly18, Gly23, Gly53 and Gly59 in Sm consensus; Fig. 2.1). The segment linking the β 4 and β 5 lies at the top of the U-shaped trough to close the protein into β -barrel-like structure.

Hfq has the same predicted secondary structure elements in the Sm1 motif region as does the Sm protein: an α helix followed by β strands. The critical residues that are required for the β 1 strand curvature (Gly29 and Gly34 in Hfq) are identical in all Hfq homologs. Interestingly, another long β strand was predicted in Hfq from Ser51 to Pro65 (in some predictions, it was two separate β strands.). Although this region could not be aligned to the Sm2 motif in the Sm protein sequence, the length of this region—referred to as Xm2 in Figure 2.1—closely matches the length of the Sm2 motif. In addition, the Sm2 and Xm2 motifs both have highly conserved residues flanked by hydrophobic

residues. There are two highly conserved residues in the middle of the Sm2 motif, Arg–Gly, and four highly conserved residues in the Xm2 motif, Tyr–Lys–His–Ala.

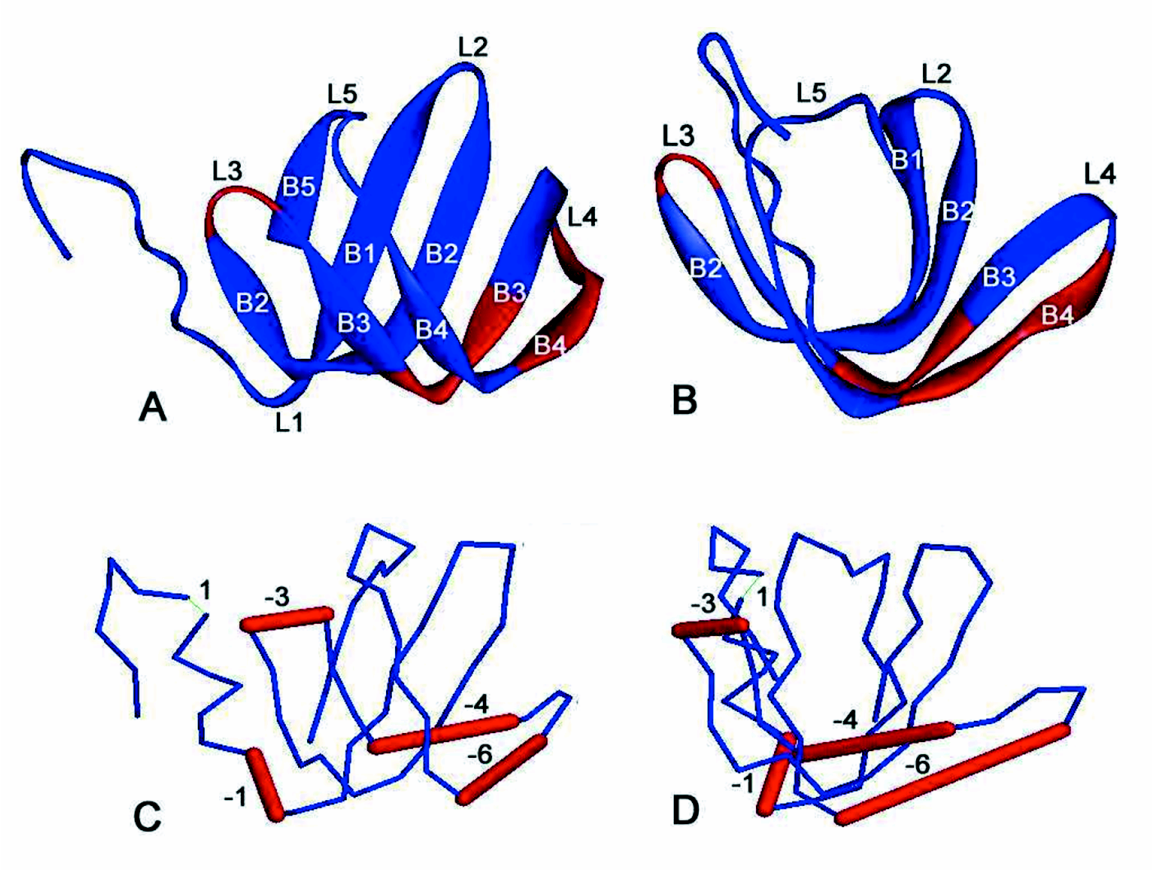


Figure 2.2 (**A** and **B**) Ribbon representations of two views of the crystal structure of an archaeal Sm protein (PDB accession number 1i81) rotated by 90˚. Images were produced by RasMol program (http://www.bernstein-plus-sons.com/software/RasMol_2.7.1). (**C** and **D**) 3D line representations of the Hfq structure predicted by the 3D-PSSM web server using the above archaeal Sm protein as template. The views shown in (C) and (D) are the same as in (A) and (B) respectively. The locations of Hfq residues that are inserted or deleted when compared with the template are represented by thin and thick bars respectively, and accompanied by numbers indicating the number of residues involved. Labels B1–B5 correspond to th β strands β 1– β 5; labels L1–L5 correspond to the loops.

The 3D structure of Hfq was predicted using comparative modeling as described in the Materials and Methods. 3D-PSSM was first employed to thread the Hfq sequence as a query against known protein structures in a fold library. The template structure that produced the highest score for the Hfq sequence was the archaeal LSm_{ar} protein from *M.thermoautotrophicum* (Collins *et al.,* 2001). The structure of the LSm_{ar} protein is shown in Figure 2.2A and B, while Figure 2.2 C and D shows the best-fit model structure of Hfq with the locations of deletions and insertions from the template structure that produced the highest score. The Hfq model is well aligned to the LSm_{ar} protein. The major difference is in the β3 and β4 region. Four and six amino acids of the template structure are absent from Hfq in the β3 and β4 strands respectively. Visually, the Hfq structure suggests that if these 10 residues are simultaneously deleted, the isolated loop 4 may be able to connect the remaining fragments of β3 and β4. The missing parts of Hfq, when compared with the LSm_{ar} protein, fall just within the highly variable region between the Sm1 and Sm2 motifs, which includes loop 4 as well as parts of the β3 and β4 strands. This suggests that the amino acid sequences which constitute a minimum Sm fold can be shortened, and may be composed of adjacent Sm1 and Sm2 motifs with no variable linker.

The SWISS-MODEL program was then employed to generate a 3D model of the Hfq protein using the archaeal LSm_{ar} protein as a template, and the information inferred from the 3D-PSSM highest scoring alignment shown in Figure 2.2 C and D. The optimized structural model of Hfq is shown in Figure 2.3 where it is compared with the structure of the LSm_{ar} protein. The features that are constant in both structures are illustrated in blue. A red ribbon designates Hfq and a green ribbon illustrates the LSm_{ar}

protein in the regions where there are differences in their structural features. The β3 and β4 strands of Hfq are shortened relative to the LSm_{ar} protein and connected by loop 4. Loop 4 changes its orientation from up and to the right for the LSm_{ar} protein to a downward direction for Hfq (Fig. 2.3 A and B). The aqua-colored ribbon shows the location of the β4 strand residues SQMVY, and the β5 strand residues AISTVV. Figure2.4 shows the β4–loop 5–β5 region in greater detail. The residues in the β5 strand of Hfq, STVVP, appear to occupy similar spatial locations as the corresponding residues of the LSm_{ar} protein, VLISP. Based on sequence alignment, amino acid characteristics and secondary structure prediction, we anticipated that the highly conserved His–Ala residues of Hfq would occupy the same 3D positions as the highly conserved Arg–Gly of an Sm protein. However, a comparison of the structural models in Figure 2.4 indicates that the Histidine and alanine residues in Hfq are shifted in their relative locations two residues downstream when compared with the Arg–Gly residues in the Sm structure.

The predicted structure of Hfq was compared with the Sm protein structure with regard to segments that may be involved in subunit interaction in the formation of multimers. Several studies indicate that Hfq forms a hexamer (Zhang *et al*., 2002; Moller *et al*., 2002; Franze de Fernandez *et al*., 1972; Kamen *et al.*, 1972). While Sm proteins form a homo- or hetero-heptamer depending on the number of distinct subunits available *in vivo*. Archea species form a heptamer composed of seven identical subunits (Toro *et al.*, 2001; Mura *et al*., 2001). Eukaryotes utilize different polypeptide chains to assemble a hetero-heptamer Sm complex (Moller *et al*., 2002; Hermann *et al.,* 1995; Kambach *et al.,* 1999). In both cases, adjacent monomers in the heptamer interact via pairing of the β4 and β5' strands (the ' indicates the adjacent subunit). Only the last five residues of the β4

strand in one Sm protein are involved in pairing with the β5' strand of the adjacent Sm protein.

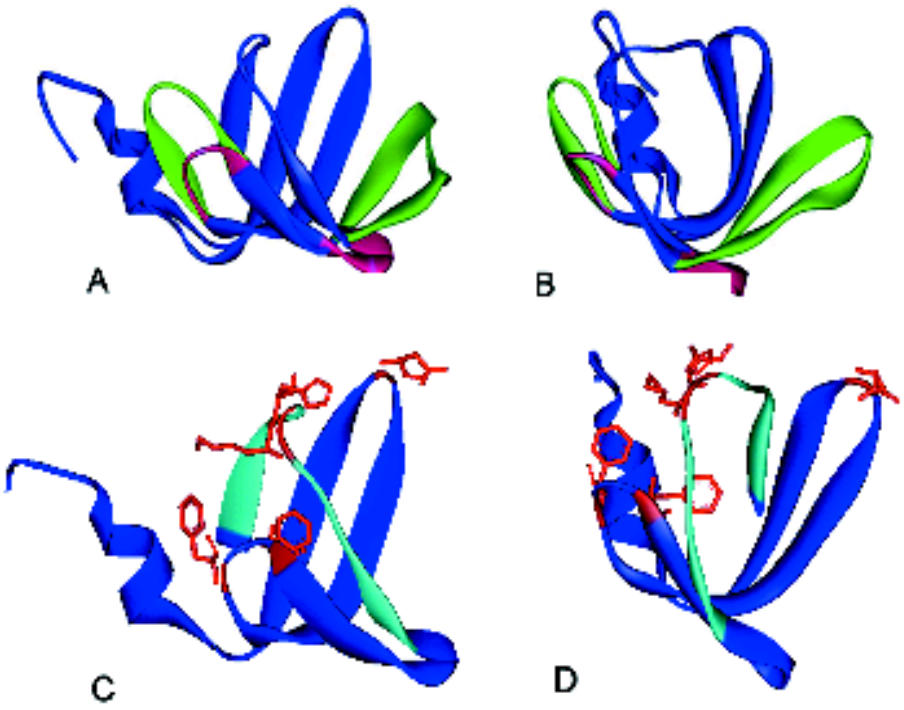


Figure 2.3 The 3D structure of Hfq generated by SWISS-MODEL program using the same archaeal LSm_{ar} protein determined to be the best template by 3D PSSM. (**A** and **B**) Front and side views of the predicted Hfq structure as well as the template Sm structure. The backbone features that are constant in both structures are illustrated in blue. Differing structural elements are shown by using a red ribbon for Hfq and a green ribbon to illustrate the LSm_{ar} protein backbone. The aqua ribbon illustrates the β4 strand residues SQMVY and β5 strand residues AISTVV. (**C** and **D**) Front and side views of the predicted Hfq model with several potential RNA-interacting residues shown in stick model representation: Lys31 in loop 2, Phe39 and Phe42 in loop 3, Lys56 and His57 in loop 5.

In the structural model of Hfq the first six residues of the β4 strand are absent when compared with the Sm protein. However, the five remaining residues that form the β4 strand and the beginning of loop 5 are located in similar positions to the residues of the

LSm protein that participate in quaternary interactions (Fig. 2.4). The residues spanning the β4–β5 strand region of the human Sm protein that are involved in pairing adjacent subunits are LVLLRGSVIVV (Kambach *et al*., 1999), while in the archaeal LSm they are TVLIRGQNIVY (Collins *et al.,* 2001). In both cases, a group of hydrophobic residues flank a positively charged Arg that is engaged in several hydrogen bonds with main chain and side chain atoms of the adjacent subunit's β5' strand. In the predicted Hfq structure, the residues spanning the β4 strand–loop– β5 strand region are SQMVYKHAISTVV. One again has hydrophobic residues flanking positively charged residues, in this case lysine and histidine. Although, as mentioned above, the His–Ala residues of Hfq are not in the same structural position as the Arg–Gly residues of the Sm protein, the similar nature of the residues in this region suggest that the predicted Hfq structure also supports multimer formation through β4–β5' strand pairing. It is worth noting that the sequence in this part of the Hfq structure, VYKHAIST, is almost completely conserved among Hfq proteins (Fig. 2.1).

The recently determined *S.aureus* Hfq structure (Schumacher *et al*., 2002) shows that β4–β5' strand interface is indeed a key part of intersubunit interactions. In this structure, H bonds occur between the highly conserved Tyr56 in β4 and Tyr63 in β5' . In the *E.coli* Hfq sequence valine occurs at position 63. This is the more dominant amino acid at this location in Hfq proteins (Fig. 2.1) and suggests that Tyr56 H-bonds with a different residue in β5' in *E.coli* Hfq or this H bond is not essential for this interface. The *S.aureus* Hfq structure also shows that contacts between -helix residues and loop L3 residues of the adjacent subunit and between side chains in β strands contribute to the dimer interface.

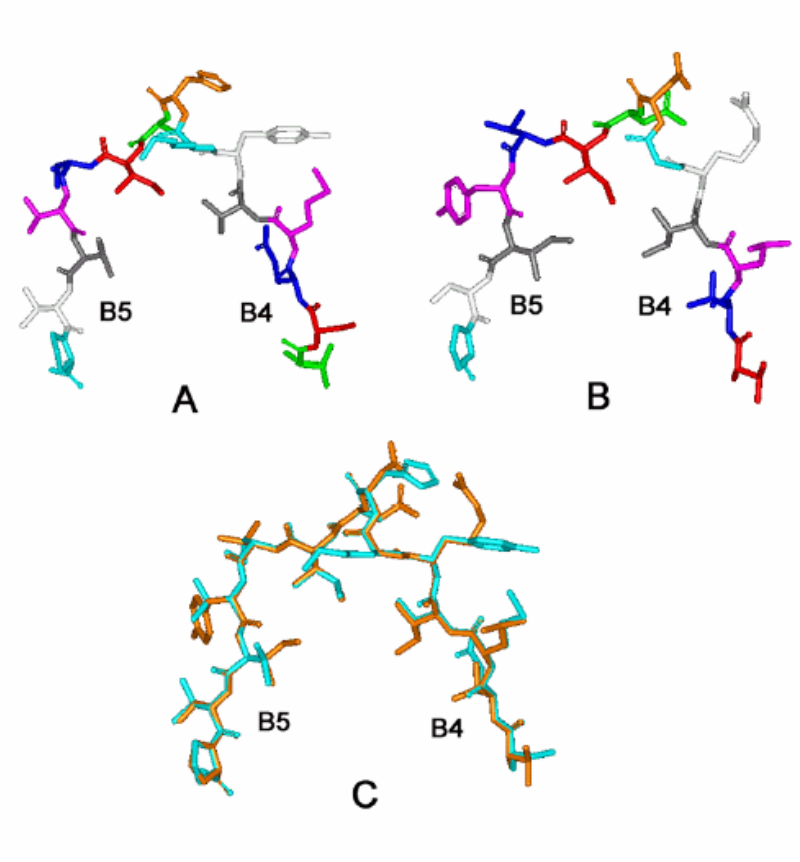


Figure 2.4 Molecular representation of the β4 and β5 strands in Hfq model (**A**) and $LSm_{\alpha\gamma}$ protein (**B**). Overlapping representation of both is shown in (**C**).

The 3D model of Hfq also provides an opportunity to consider its potential sites of interaction with RNA. The RNA determinants important for Sm core assembly appear to be complex. One prerequisite for an RNA to be bound by an Sm protein heptamer is an 'Sm site element', a 7–10 nt single-stranded segment that has the consensus sequence $PuAU_{3-6}GPu$ usually flanked by stem–loop structures (Branlant *et al*., 1982; Raker *et al*., 1999). *In vitro* analysis with an RNA oligonucleotide consisting of a minimal Sm site element revealed that the 5' adenosine of the element plays a critical role in the heptamer's association, while the uridine bases and the 2' hydroxyl groups collectively provide a binding determinant (Raker *et al*., 1999; Hartmuth *et al*., 1998).

In human snRNP core, several Sm proteins were shown to interact with the uridine stretch of the Sm site element by UV cross-linking experiments. The most efficient cross-links were observed for the G and B/B' proteins, which are linked to the first and third uridines of the Sm site element respectively (Urlaub *et al.*, 2001). The residues (His37 for B/B', Phe37 for G) involved in contacting the RNA are located at equivalent regions in both proteins, namely in loop L3 of the Sm1 motif. In contrast, crystal structure of the archaeal SmAP protein suggests that residues in other loops (Arg29 in L2, Asp57 in L4 and Glu71 in L5) are more likely to interact with the RNA Sm site element (Mura *et al.,* 2001). All four of these loops jut into the inner ring or pore of the doughnut-shaped heptamer (Fig. 2). The corresponding regions of Hfq, which by analogy would be expected to be oriented toward the inner ring of the hexamer, also have conserved residues (for the *E.coli* sequence: Lys31 in loop 2, Phe39 and Phe42 in loop 3, Lys56 and His57 in loop 5). The location of some of these residues is illustrated in Figure 3. The Phe39 and Phe42 in loop 3 and Lys56–His57 in loop 5 are almost 100% conserved among different bacterial species examined, implying they have critical roles in structure and function.

Hfq has been shown to be an essential participant in facilitating the interaction of some small riboregulator RNAs, such as DsrA (Sledjeski *et al*., 2001) and Spot42 (Moller *et al.,* 2002), with their target mRNAs. It was proposed that the role of Hfq might be analogous to Rop, in which two phenylalanines intercalate into base pairs and facilitate the pairing of two RNA molecules (Predki *et al.,* 1995). If Hfq functions in this way, the highly conserved Phe42 in loop 3 and the nearby Phe39 are candidates for this role (Fig. 2.3 C and D).

Several of the above predictions were verified in the crystal structure of the *S.aureus* Hfq–RNA complex (Schumacher *et al*., 2002). In this structure, the backbone of the oligoribonucleotide 5'-AUUUUG-3' was found to form a circular conformation as it bound to an electropositive patch around one face of the pore of the hexameric Hfq. Residues in the Sm1 and Sm2 motifs of adjacent subunits are utilized to build six nucleotide-binding pockets. There are no intramolecular base stacking interactions within the RNA as the bases are splayed out, fitting into the individual binding pockets. Each base is sandwiched between two Tyr42 side chains from adjacent subunits. The presence of Phe42 instead of Tyr42, which occurs for most Hfq sequences, may be able to serve the same function for the nucleotide-binding pockets. The highly conserved Lys 57–His58 motif located in loop 5 and facing the pore also contacts the RNA. Lys57 (shifted by one amino acid in *S.aureus* due to an extra residue relative to the *E.coli* sequence) H-bonds with uracil and His58 makes contacts with the phosphate oxygens of one nucleotide as well as the ribose O2' hydroxyl of the adjacent nucleotide.

Phyletic distribution of the Hfq protein

In order to determine the pervasiveness of the Hfq gene and related Sm proteins in bacteria and explore its evolution, a BLAST search was conducted against bacterial genomes available at the NCBI database. Fifty-eight completed and 82 unfinished bacterial genomes were examined. Approximately half of the bacterial genomes contain at least one gene that codes for an Hfq protein based on strong amino acid sequence similarity to the *E.coli* Hfq sequence, sequence length and amino acid conservation pattern. The presence and absence of the Hfq protein in particular species closely follows

recently redefined major bacterial clades (Wolf *et al*., 2001; Brochier *et al*., 2002), as shown in Table 2.1.

Hfq was missing from three high-level bacterial clades: Chlamydia-Spirochaetes, Actinomycetes-Deinococcus- Cyanobacteria and Green sulfur bacteria-Cytophagales. However, it is present in the most deeply branched Thermotogales-Aquificales. The Hfq protein is present in all alpha, beta, gamma and delta proteobacteria except for those that have experienced a massive genome reduction due to their parasitic lifestyle, e.g. *Buchnera* sp. (Gil *et al.*, 2002), *Rickettsia prowazekii* (Anderson *et al*., 1998). This type of phyletic distribution suggests two possible scenarios for the evolution of the Hfq protein. First, Hfq might be an ancient protein, which was lost early in evolution by major clades and retained only by the lineage leading to proteobacteria. Second, the Hfq protein might have evolved late in evolution during the separation of the proteobacterial clade and was transferred laterally to several species outside the proteobacteria.

A phylogenetic tree built from the multiple alignment of the Hfq protein sequences is shown in Figure 2.5. The tree has a topology expected from Table 2.1. Three major clades of alpha proteobacteria, beta/gamma proteobacteria and low GC Gram-positive bacteria are well defined and supported by bootstrap analysis. A comparison of Hfq sequences from bacteria in the three groups illustrated in Figure 2.5 is given in Figure 2.6. The results reveal three positions, at the borders of loop 2 and loop 3, and in loop 4 of the predicted Hfq structure that have residues characteristic to each group. For the low GC gram-positive bacteria, the dominant residues corresponding to the *E.coli* sequence positions 30, 43 and 50 are phenylalanine, tyrosine and lysine respectively. The alpha proteobacteria are dominated at these positions by valine, cysteine, and histidine or

glutamine, while the beta/gamma proteobacteria universally have isoleucine, valine and valine at these locations. The positions are highlighted in Figure 2.6. These residues may provide some specificity in the interactions of Hfq with RNA or the interactions governing subunit oligomerization.

Since the above analysis is based on the assumption that the model structure for the *E.coli* Hfq is appropriate for other Hfq proteins, it is worth noting that our predicted structure was in excellent agreement with the crystal structure of the *S.aureus* Hfq (Schumacher *et al.,* 2002) as well as *E coli* Hfq (Sauter *et al.,* 2003) and *Pseudomonas aeruginosa* Hfq (Nikulin *et al.,* 2005). We also note that four independent algorithms which utilize a single amino acid sequence as a query predict secondary structures for *S.aureus* Hfq and other relatively distant Hfq sequences (e.g. *Geobacter sulfureducens* Hfq) that closely fit the LSm and consensus Hfq structures (data not shown).

Another outcome from the microbial genome search worth noting was that 6 of the 140 eubacterial genomes examined contained two distinct copies of an Hfq protein coding sequence (Table 1). Duplicated *hfq* genes are always found within the same clade on the phylogenetic tree as original copies (Fig. 5) indicating the likelihood of paralogous relationships over lateral gene transfer. Two Hfq sequences are found within a single 193-residue protein of the bacterium *Novosphingobium aromaticivorans*, which reinforces the notion that Hfq is a subject of relatively frequent gene duplication events. Twenty-three residues separate the two distinct 59 residue Hfq motifs (E values of 3×10^{-14} and 2×10^{-10}) in the *N.aromaticivorans* protein. This protein may code for a heterodimeric version of an Hfq structural unit similar to the heterodimers observed for the eukaryotic Sm proteins. Our phylogenetic analysis produced no evidence for lateral transfer of Hfq. This is

consistent with the proposal that the bacterial (Hfq) and archaeal/eukaryotic (Sm and LSm) versions of this important RNA-binding protein shared a common ancestor prior to the separation of bacteria and archaea–eukarya. Gene loss appears to be a major driving force in the evolution of Hfq.

Table 2.1 Presence and absence of Hfq from BLAST search of bacterial genomes

Phyla	Species with Hfq	Species without Hfq
Thermotogales-Aquificales	*Aquifex aeolicus*	
	Thermotoga maritima	
Chlamydia-Spirochetes		*Chlamydia muridarum*
		Chlamydia trachomatis
		Chlamydophlia pneumoniae
		Borrelia burgdorferi
		Treponema pallidum
Green sulfur-Cytophagales		*Cytophaga hutchinsonii*
Actinomycetes- Deinococcales		*Deinococcus radiodurans*
Cyanobacteria		*Mycobacterium leprae*
		Streptomyces coelicolor
		Thermobifida fusca
		Nostoc sp.
		Nostoc punctiforme
		Synechocystis sp.
		Prochlorococcus marinus
		Synechococcus sp.
Low GC Gram-positive	*Bacillus halodurans*	*Mycoplasma genitalium*
	Bacillus subtilus	*Mycoplasma pneumoniae*
	Clostridium acetobutylicum	*Mycoplasma pulmonis*
	Clostridium perfringens	*Ureaplasma urealyticum*
	Listeria innocua	*Lactococcus lactis*
	Listeria monocytogenes	*Streptococcus pneumoniae*
	Staphylococcus aureus	*Streptococcus pyogenes*
	Bacillus anthracis(2)	*Enterococcus faecium*
ε-proteobacteria		*Helicobacter pylori*
		Campylobacter jejuni
α-proteobacteria	*Agrobacterium tumefaciens*	
	Brucella melitensis	
	Caulobacter crescentus	*Rickettsia conorii*
	Mesorhizobium loti	*Rickettsia prowazekii*
	Sinorhizobium meliloti	
	Magnetococcus sp.	
	Rhodopseudomonas palustris	
	Rhodobacter sphaeroides	
	Magnetospirillum magnetotacicum(2)	
	Novosphingobium aromaticivorans(2)	
β/γ-proteobacteria	*Neisseria meningitidis*	*Buchnera sp.*
	Ralstonia solanocearum	
	Burkholderia fungorum (2)	
	Burkholderia mallei(2)	
	Burkholderia psudomallei(2)	
	Haemophilus influenza	
	Pasteurella multocida	
	Xylella fastidiosa	
	Nitrosomonas europea	
	Pseudomonas aeruginosa	
	Pseudomonas fluorescens	
	Vibrio cholerae	
	Yersinia pestis	
	Salmonella enterica	
	Escherichia coli	
	Shewanlla putrefaciens	

Species with two copies of Hfq sequences per genome are followed by (2)

Figure 2.5 Unrooted neighbor-joining tree inferred by analysis of Hfq protein sequences. Sequences were aligned using CLUSTAL program and all positions with gaps were excluded from the analysis. Bootstrap values of >600 are displayed at deep nodes only. Color code: green, low GC Gram-positive bacteria; red, alpha proteobacteria; purple, beta proteobacteria; blue, gamma proteobacteria; orange, delta proteobacteria. Aquifecales-Thermatogales and unclassified *Magnetococcus* are shown in black.

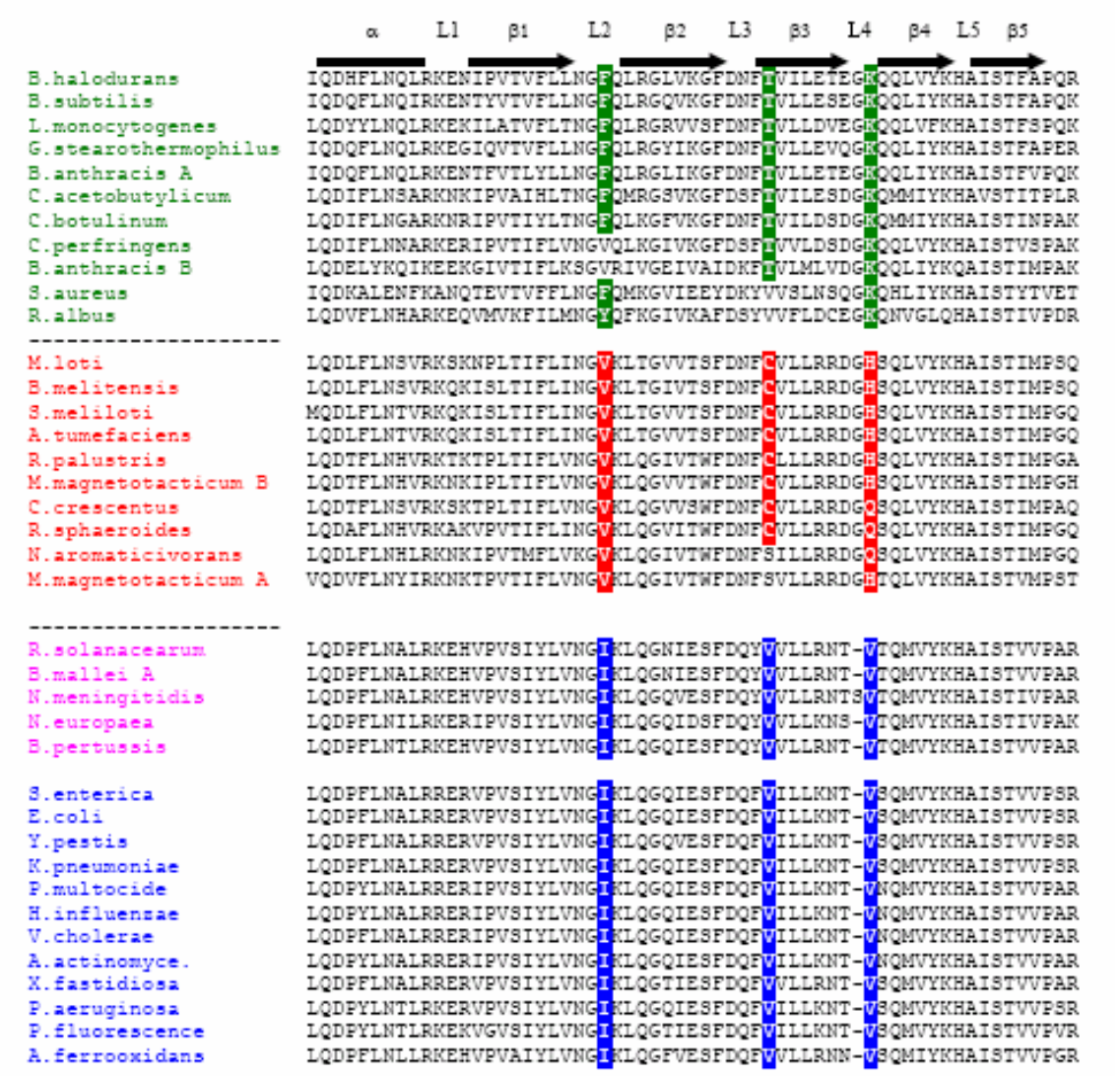


Figure 2.6 Conserved amino acid residues specific to Hfq proteins from major bacterial groups defined by phylogenetic analysis. Multiple alignment of Hfq sequences is subdivided according to bacterial groups inferred from the tree shown in Figure 5. Positions where amino acid conservation are group specific are shown.

References

Achsel,T., Stark,H. and Luhrmann,R. (2001) The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. *Proc. Natl Acad. Sci. USA*, **98**, 3685–3689.

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Andersson,S.G., Zomorodipour,A., Andersson,J.O., Sicheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K., Eriksson,A.S., Winkler,H.H. and Kurland,C.G. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, **396**, 133–140.

Branlant,C., Krol,A., Ebel,J.P., Lazar,E., Haendler,B. and Jacob,M. (1982) U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *EMBO J.*, **1**, 1259–1265.

Brochier,C., Bapteste,E., Moreira,D. and Philippe,H. (2002) Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet.*, **18**, 1–5.

Collins,B.M., Harrop,S.J., Kornfeld,G.D., Dawes,I.W., Curmi,P.M. and Mabbutt,B.C. (2001) Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs. *J. Mol. Biol.*, **309**, 915–923.

Cuff,J.A., Clamp,M.E., Siddiqui,A.S., Finlay,M. and Barton,G.J. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics*, **14**, 892–893.

DelVecchio,V.G., Kapatral,V., Redkar,R.J., Patra,G., Mujer,C., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A., Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A., Mazur,M., Goltsman,E., Selkov,E., Elzer,P.H., Hagius,S., O'Callaghan,D., Letesson,J.J., Haselkorn,R., Kyrpides,N. and Overbeek,R. (2002) The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. *Proc. Natl Acad. Sci. USA*, **99**, 1–3.

Fischer,U., Sumpter,V., Sekine,M., Satoh,T. and Luhrmann,R. (1993) Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *EMBO J.*, **12**, 573–583.

Franze de Fernandez,M.T., Hayward,W.S. and August,J.T. (1972) Bacterial proteins required for replication of phage Q ribonucleic acid. Purification and properties of host factor I, a ribonucleic acid-binding protein. *J. Biol. Chem.*, **247**, 824–831.

Gil,R., Sabater-Munoz,B., Latorre,A., Silva,F.J. and Moya,A. (2002) Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. *Proc. Natl Acad. Sci. USA*, **99**, 4454–4458.

Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.

Hartmuth,K., Raker,V.A., Huber,J., Branlant,C. and Luhrmann,R. (1998) An unusual chemical reactivity of Sm site adenosines strongly correlates with proper assembly of core U snRNP particles. *J. Mol. Biol.*, **285**, 133–147.

Hermann,H., Fabrizio,P., Raker,V.A., Foulaki,K., Horning,H., Brahms,H. and Luhrmann,R. (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J.*, **14**, 2076–2088.

Kambach,C., Walke,S., Young,R., Avis,J.M., de la Fortelle,E., Raker,V.A., Luhrmann,R., Li,J. and Nagai,K. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, **96**, 375–387.

Kamen,R., Kondo,M., Romer,W. and Weissmann,C. (1972) Reconstitution of Q replicase lacking subunit with protein-synthesis-interference factor. *Eur. J. Biochem.*, **31**, 44–51.

Kelley,L.A., MacCallum,R.M. and Sternberg,M.J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, **299**, 499–520.

Lerner,M.R. and Steitz,J.A. (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl Acad. Sci. USA*, **76**, 5495–5499.

Luhrmann,R., Kastner,B. and Bach,M. (1990) Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim. Biophys. Acta*, **1087**, 265–292.

Moller,T., Franch,T., Hajrup,P., Keene,D.R., Bahinger,H.P., Brennan,R.G. and Valentin-Hansen,P. (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol. Cell*, **9**, 23–30.

Mura,C., Cascio,D., Sawaya,M.R. and Eisenberg,D.S. (2001) The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. *Proc. Natl Acad. Sci. USA*, **98**, 5532–5537.

Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, Moll I, Kachalova G, Yokoyama S, Vassylyev D, Garber M, Nikonov S. (2005) Structure of Pseudomonas aeruginosa Hfq protein. Acta Crystallogr D Biol Crystallogr, **61**:141-6.

Predki,P.F., Nayak,M., Gottlieb,M.B. and Regan,L. (1995) Dissecting RNA-protein interactions: RNA-RNA recognition by Rop. *Cell*, **80**, 41–50.

Raker,V.A., Hartmuth,K., Kastner,B. and Luhrmann,R. (1999) Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol. Cell. Biol.*, **19**, 6554–6565.

Salgado-Garrido,J., Bragado-Nilsson,E., Kandels-Lewis,S. and Seraphin,B. (1999) Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J.*, **18**, 3451–3462.

Sauter C, Basquin J, Suck D. (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. Nucleic Acids Res. **31**(14):4091-4098.

Schumacher,M.A., Pearson,R.F., Moller,T., Valentin-Hansen,P. and Brennan,R.G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.*, **21**, 3546–3556.

Seraphin,B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.*, **14**, 2089–2098.

Sledjeski,D., Whitman,C. and Zhang,A. (2001) Hfq is necessary for regulation by the untranslated RNA DsrA. *J. Bacteriol.*, **183**, 1997–2005.

Sonnleitner,E., Moll,I. and Blasi,U. (2002) Functional replacement of the *Escherichia coli* hfq gene by the homologue of *Pseudomonas aeruginosa*. *Microbiology*, **148**, 883–891.

Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

Toro,I., Thore,S., Mayer,C., Basquin,J., Seraphin,B. and Suck,D. (2001) RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J.*, **20**, 2293–2303.

Urlaub,H., Raker,V.A., Kostka,S. and Luhrmann,R. (2001) Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.*, **20**, 187–196.

Walke,S., Bragado-Nilsson,E., Seraphin,B. and Nagai,K. (2001) Stoichiometry of the Sm proteins in yeast spliceosomal snRNPs supports the heptamer ring model of the core domain. *J. Mol. Biol.*, **308**, 49–58.

Wolf,Y.I., Rogozin,I.B., Grishin,N.V., Tatusoy,R.L. and Koonin,E.V. (2001) Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol. Biol.*, **1**, 8–22.

Zhang,A., Wassarman,K.M., Ortega,J., Steven,A.C. and Storz,G. (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell*, **9**, 11–22.

CHAPTER III

LOCATING THE SURFACE RESIDUES INVOLVED IN RNA BINDING

Introduction

While structural details of Hfq have been revealed and its multiple functional roles in regulating gene expression demonstrated, the molecular mechanisms governing its function remain unclear. As described in chapter I, two parallel but nonexclusive models have been proposed to explain how Hfq promotes intermolecular base-pairing. In the first model, Hfq acts explicitly as an RNA chaperone, partially unfolding one or both RNAs (Storz *et al*., 2004). Evidence indicates Hfq may act as a RNA chaperone and modulate RNA structure (Moll *et al.*, 2003b). In the second model, Hfq binds both RNAs, increasing their local concentration to induce the interaction. The latter role is supported by an earlier finding that *E coli* Hfq can bring together distant segments of Qβ RNA to form a looped structure (Miranda *et al.*, 1997) and results that indicate that this Hfq can simultaneously bind to more than one RNA (Brescia *et al.,* 2003; Altuvia *et al.,* 1998; Mikulecky *et al.,* 2004). Addition of poly A did not compete with DsrA RNA for Hfq binding and produced a supershifted gel complex suggestive of a ternary complex. Similarly, a supershifted band was observed when Hfq was examined with OxyS RNA and its *fhlA* mRNA target, and spot42 RNA and its *galK* mRNA target. Coimmunoprecipitation assays also implied formation of a tertiary complex.

A recent study (Mikulecky *et al.,* 2004) indicated that *E coli* Hfq has distinct interaction surfaces for DsrA RNA and $(A)_n$ sequences. Several Hfq residues critical for RNA binding were identified by site-directed mutagenesis. Mutant Hfqs were grouped into three categories: those with mutations affecting the central cavity (Y55A, K56A and H57A), the proximal surface (Q8A, D9A, D40A, Q41A, F42A and I59A) and the distal surface (Y25X, I30X, Q53A and S60A). Mutations at Y25 and I30 decreased Hfq binding to A_{27} 5 to 10 fold, but did not affect binding to DsrA RNA. Only two cavity mutants Y55A and K56A showed defects in binding DsrA RNA. Interestingly, no other residues examined on the proximal surface were found to affect DsrA binding. Since the U-rich stretch of DsrA is protected by Hfq binding, and other studies imply this sequence binds to the proximal surface, other contacts between DsrA and the proximal surface may exist.

To address this question, a search for other potential RNA binding residues was carried out using comparative sequence and structural analysis of E coli Hfq. Based on this analysis we examined the effect of mutations of several Hfq surface residues on binding two RNAs, domain II of DsrA RNA and A_{18} by using fluorescence quenching, fluorescence anisotropy and gel shift assay. Several potential RNA binding residues on both surfaces of *E coli* Hfq were identified and an amino-acid covariance was observed for L12 and F39 in the proximal surface of Hfq hexamer.

Materials and Methods

General Material

ER2566 *E. coli* strain was purchased from New England Biolabs Inc (Beverly, MA). Competent cells were made by general methods (Sambrook and Russll 2001). *SapI* and *Sma I* restriction endonucleases, Vent DNA polymerase and T4 ligase were purchased from New England Biolabs Inc (Beverly, MA). Bio-Rad Protein Assay reagent was purchased from Bio-Rad Laboratories (Hercules, CA). The Wizard PCR Prep DNA purification System, The Wizard Plus Minipreps DNA Purification System was purchased from Promega Corporation (Madison, WI). The QuickChange™ Site-Directed Mutagenesis Kit was purchased from Stratagen (La Jolla, CA). DNA oligonucleotides were purchased from the Integrate DNA Technology (Coralville, IA). The fluorescein labeled or non-labeled RNA oligomers were purchased from Dharmacon Research Inc (Chicago, IL). Isopropyl-β-D-thiogalactopyranoside (IPTG), dithiothreitol (DTT), ampicillin, DEAE cellulose were purchased from Sigma Chemical (St. Louis, MO). Tris-HCl, Tris-base, boric acid, ethylene-diamine-tetracetic acid (EDTA), LB-Broth (Luria-bertani) Miller, LB-Agar, Sodium Chlorate, Sodium hydroxide, SDS was purchased from Fisher Scientific (Fair Lawn, NJ). Centricon-3 and Centriprep-10 were purchased from the Millipore. A Cary-1E UV-Visible Spectrophotometer was used for absorption spectroscopy. A Bio-Rad MJ mini Thermo-Cycler was used for amplification of the *hfq* gene from *E.coli* genomic DNA and the temperature cycling component of the QuickChange™ Site-Directed Mutagenesis Kit. A Jasco810 CD spectrophotometer was used for the CD spectroscopy. An ISS fluorimeter was used to do the fluorescence quenching and fluorescence anisotropy studies. A Fuji FLA-3000 Imager was used for imaging the gel mobility shift assay. An alpha Gel imaging System was used for the

SDS-PAGE and agarose gel. DNA sequencing was performed at the DNA Sequencing Core Facility at School of Biology, Georgia Institute of Technology (Atlanta, GA).

Cloning of *E coli* Hfq

The gene encoding *E coli* Hfq was amplified from E coli genomic DNA by polymerase chain reaction (PCR). The 5' primer (5'-GGTGGTTGCTTCCAAC ATGGCTAAGGGGCAATCTTTACAAGATC-3') was designed to contain the 5' portion of Hfq as well as harbor a *Sap I* restriction site 5' to the beginning of the coding sequence. The 3' primer (5'-TTATTCGGTTTCTTCGCTGT CCTGTT-3') was designed to contain the 3' portion of Hfq. The PCR reaction mixture can be found in Table 3.1. The reaction mixture was subject to thermal cycling indicated in Figure 3.1. The amplified DNA fragments were purified using the Wizard PCR Preps DNA purification System as per the protocol of the manufacturer (Promega).

Table 3.1 *hfq* gene PCR reaction

10X PCR buffer	5 µl
$MgSO_4$ (100mM)	2 µl
Hfq primer 1 (5µM)	5 µl
Hfq primer 2 (5µM)	5 µl
dNTP (5mM)	2 µl
Vent polymerase (2 unit/µl)	1 µl
E coli genome (300ng/µl)	2 µl
Water	28 µl
Total	50 µl

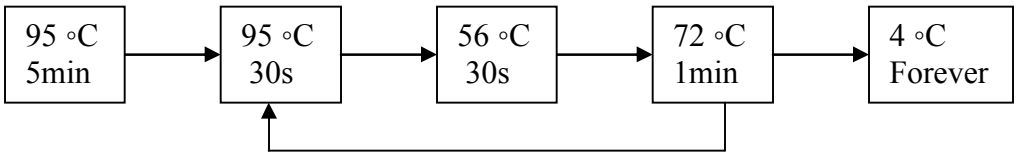
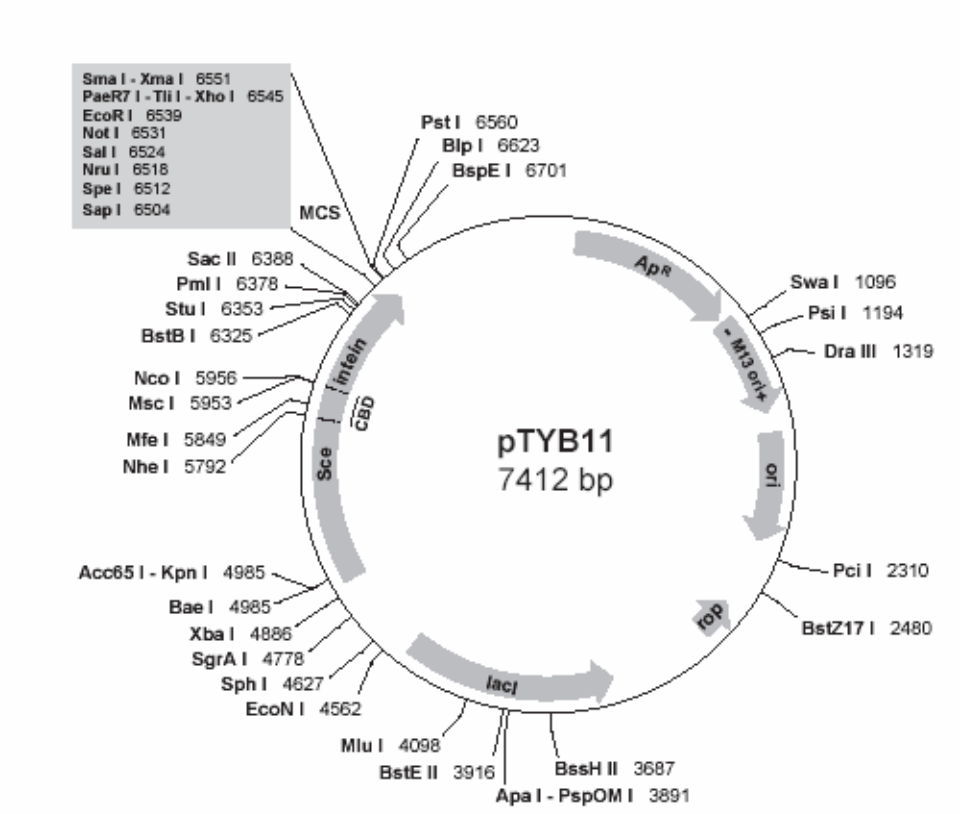


Figure3.1 Thermal cycling of PCR reaction

The pTYB11 vector (*New England Biolab*) was used for cloning and expression of the recombinant Hfq protein. pTYB11 (7,414 bp) is an N-terminal fusion vector in which the N-terminus of the target protein is fused to the intein tag used for purification (Figure 3.2). The pTYB11 vector was digested with *SmaI* and *SapI* to produce a linear plasmid with appropriate sticky and blunt ends in the multiple clone site (MCS). The purified DNA fragment from the PCR amplification (above) was also subjected to the same restriction enzyme digestion. Both reaction mixtures are loaded on a 1% low-melting agarose gel. The gel slices containing the digested gene fragment and digested pTYB11 vector were cut out and purified by the QIAquick Gel Extraction Kit according to the protocol of the manufacturer (QIAGEN). The resulting two DNA molecues were mixed at an appropriate ratio (the molar ratio of the linear pTYB11 plasmid to the *hfq* PCR product is roughly equal to 1:5) and ligated using T4 DNA ligase at 16 ˚C for 16 hours in a buffer provided by the manufacturer. The ligation mixture was used to transform ER2566 *E. coli* cells rendered competent using $CaCl_2$ from general methods described above. The randomly picked colonies were grown up in LB and subject to plasmid purification using the Wizard Plus Minipreps DNA Purification System kit as per the protocol of the manufacturer (Promega). The isolated plasmids were then digested by

a set of appropriate restriction enzymes (Apa I and EcoR I) and those containing the correct target gene insert are identified. These clones were further confirmed by DNA sequencing. The pTYB11 vector harboring the wild type *hfq* will be referred to as pwtHfq.

A



B

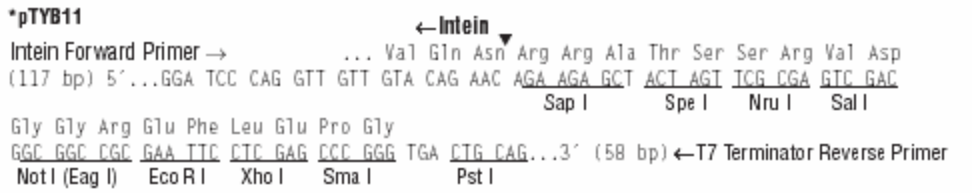


Figure 3.2. **A.** pTYB11 plasmid. **B.** The multiple cloning site of pTYB11. The pTYB11 vector use a T7/lac promoter to provide stringent control of the fusion gene expression. The vector carry its own copy of the *lac I* gene endcoding the lac repressor. Bind of the lac repressor to the lac operator sequence immediately downstream of the T7 promoter suppresses basal expression of the fusion gene in the absence of IPTG induction. The vector carries the Amp^r gene marker, which conveys ampicillin resistance to the host strain.

Site-directed Mutagenesis of Hfq

Site-directed mutagenesis of *E. coli* Hfq was performed using QuickChange™ Site-Directed Mutagenesis Kit. The basic procedure (see Figure 3.3 for a schematic representation of the procedure) utilizes a supercoiled double-stranded DNA (dsDNA) vector with an insert of interest and two synthetic oligonucleotide primers containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector are extended during temperature cycling by *PfuTurbo* DNA polymerase. Incorporation of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling, the product is treated with *Dpn* I. The *Dpn* I endonuclease (target sequence: 5'-Gm^{6}ATC-3') is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. DNA isolated from almost all *E.coli* strains is dam methylated and therefore susceptible to *Dpn* I digestion. The nicked vector DNA containing the desired mutations is then transformed into XL1-Blue supercompetent cells and grown in LB-Amp plate. Greater than 80% of the colonies should contain the mutation. To generate the mutant *hfq* plasmid, a reaction mixture (see Table3.2) contained the pwtHfq parent plasmid, appropriate buffer, dNTPs and complimentary oligonucleotide mutant primers (see Table 3.3) was made. The mixture was then subject to a thermal cycling using the program shown in Figure3.3. Next, the products were incubated with provided *Dpn* I restriction endocuclease to digest the non-mutated methylated parental template DNA. XL1-Blue super-competent cells were transformed with the mixture containing circular, nicked dsDNA. Colonies were selected and grown overnight in LB broth (100μg/ml ampicillin). Plasmid preparation using Wizard Plus

Minipreps DNA purification System was used to isolate putative mutant plasmids for sequencing to confirm proper mutagenesis.

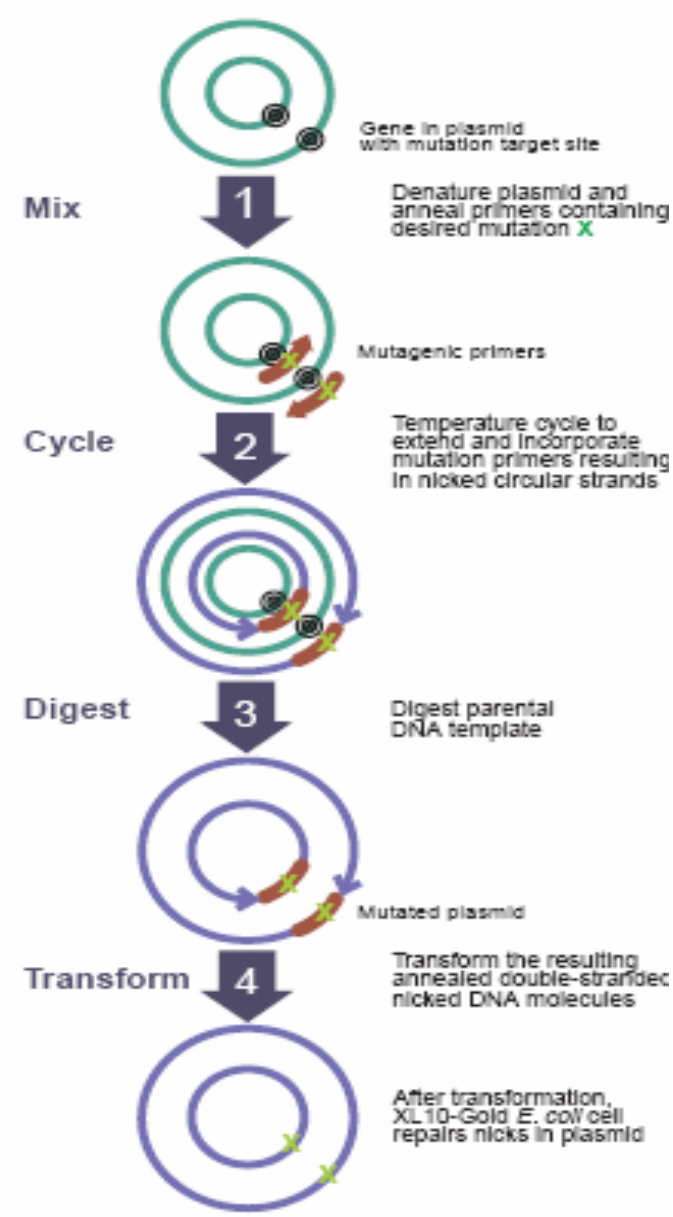


Figure 3.3 Overview of the QuickChange site-directed mutagenesis method (from manual)

Table 3.2 PCR reaction for site-directed mutagensis

10X PCR buffer	5 µl
Mutation primer 1 (50 ng/µl)	3 µl
Mutation primer 2 (50 ng/µl)	3 µl
dNTP (5mM)	1 µl
PfuTurbo polymerase (2.5 U/µl)	1 µl
pwtHfq (10ng/µl)	2 µl
Water	35 µl
Total	50 µl

Table 3.3 Primers used for site-directed mutagenesis PCR

Mutation	Primer Sequences
F39A	5'-GGGCAAATCGAGTCTGCTGATCAGTTCGTGATCCTGTTG-3'
	5'-CAACAGGATCACGAACTGATCAGCAGACTCGAT TTGCCC-3'
F42A	5’-CGAGTCTTTTGATCAGGCCGTGATCCTGTTGAAAAACACGG-3’
	5’-CCGTGTTTTTCAACAGGATCACGGCCTGATCAAAAGACTCG-3’
F42A/F39A	5’-CGAGTCTGCTGATCAGGCCGTGATCCTGTTGAAAAACAC-3’
	5’ GTGTTTTTCAACAGGATCACGGCCTGATCAGCAGACTCG-3’
L12F/F39A	5’-CAATCTTTACAAGATCCGTTCTTCAACGCACTGCGTCGGG-3’
	5’-CCCGACGCAGTGCGTTGAAGAACGGATCTTGTAAAGATTG-3’
Q8A	5’-GGCTAAGGGGCAATC TTTAGCAGATCCGTTCCTGAACGC-3’
	5’-GCGTTCAGGAACGGATCTGCTAAAGATTGCCCCTTAGCC-3’
R16A	5’-5CCTGAACGCACTGGCTCGGGAACGTGTTCC-3’
	5’-GGAACACGTTCCCGAGCCAGTGCGTTCAGG-3’
Y25A	5’-GAACGTGTTCCAGTTTCTATTGCTTTGGTGAATGGTATTAAGCTG-3’
	5’CAGCTTAATACCATTCACCAAAGCAATAGAAACTGGAACACGTTC3’
K31A	5’-TTTATTTGGTGAATGGTATTGCGCTGCAAGGGCAAATCGAGTC-3’
	5’- TTTATTTGGTGAATGGTATTGCGCTGCAAGGGCAAATCGAGTC-3’

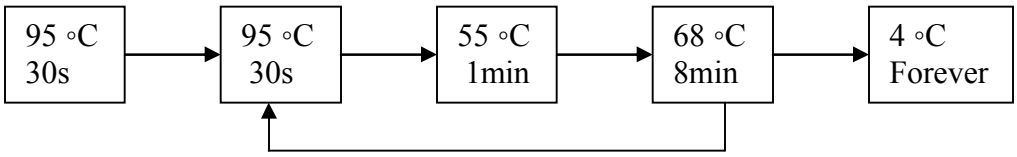


Figure3.4 Thermal cycling of PCR reaction for site-directed mutagenesis

Overexpress and Purification of Hfq Protein

Expression of wild type and mutant Hfq proteins was accomplished by transformation of ER2566 *E coli* cells with the purified plasmids. Transformed cells were maintained on LB-agar plates containing 100 μg/ml ampicillin and used to first inoculate a 10 ml preculture and than 1 liter of LB broth with 100 μg/ml ampicillin. The culture was incubated in an air shaker at 37 ◦C at 225 RPM until the absorbance A_{600} reaches 0.6-0.8 OD. IPTG was added to a final concentration of 0.5 mM before the culture was then transferred to a 15 ˚C air shaker. This allows accumulation of sufficient T7 RNA polymerase for protein expression. Induction at 15 ˚C should be conducted overnight. The cells from the above culture are spun down at 5000g for 10 minutes at 4 ˚C. After discarding the supernatant, the cell pellet can be stored at -20 ˚C.

The proteins were purified using the IMPACT (Intein Mediated Purification with an Affinity Chitin-binding Tag) protein purification system which utilizes the inducible self-cleavage activity of a protein splicing element (termed intein) to separate the target protein from the affinity tag. The IMPACT-CN system utilizes an intein (454 amino acid residues) from the *Saccharomyces cerevisiae* VMA1 gene. A target protein is fused to a self-cleavable intein tag in which a chitin binding domain allows affinity purification of

the fusion precursor on a chitin column. In the presence of thiols such as DTT, the intein undergoes specific self-cleavage which releases the target protein from the chitin-bound intein tag resulting in a single-column purification of the target protein (Figure 3.5). The chitin-binding domain in the intein tag has an extremely high affinity for the chitin beads, which allows efficient recovery of the fusion protein from the crude cell extract. 18 ml of chitin beads was used for one liter culture. The chitin beads was equilibrated at 4 ˚C with 10 volumes of the Column Buffer (0.5 M NaCl, 20mM Tirs-HCl; pH 8, 1mM EDTA) prior to the loading of the crude cell extracts. The cell pellet from one liter culture is resuspended in 40 ml ice-cold cell lysis buffer (0.5 M NaCl, 20mM Tirs-HCl; pH 8.0, 1mM EDTA, 0.1% Triton X-100) and broken by French press 4 times. The clarified cell extract (supernatant) is obtained by centrifugation at 20,000g for 30 minutes and mixed with the column buffer to bring up the volume to 100ml which is then slowly loaded onto the chitin column at a flow rate no faster than 0.5-1 ml/min. The column was then washed by 15 bed volumes of the column buffer (1.0M NaCl, 20mM Tirs-HCl; pH 8.0, 1mM EDTA) with a higher flow rate (2ml/min) to remove any non-specifically bound proteins. Induction of the on-column self-cleavage is conducted by quickly flushing the column with 3 bed volumes of the Cleavage buffer (0.5M NaCl, 20mM Tirs-HCl; pH 8, 1mM EDTA) containing 40 mM of DTT. After the quick flush, the flow in the column is stopped, and the column left at room temperature for 48 hours. After the incubation of the cleavage reaction, the target Hfq protein is released from the intein tag by washing the column with 60 ml column buffer (0.5 M NaCl, 20mM Tirs-HCl; pH 8, 1mM EDTA).

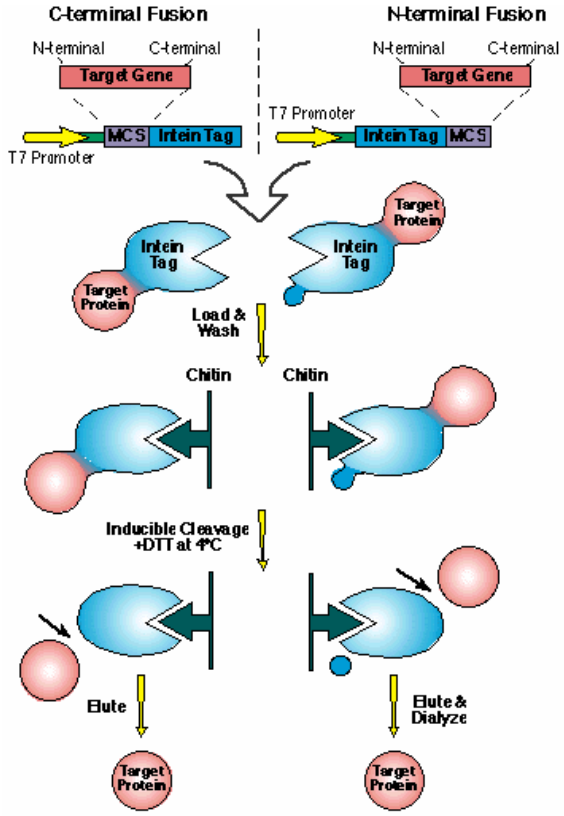


Figure3.5 A schematic illustration of the IMPACT-CN System (from IMPACT manual)

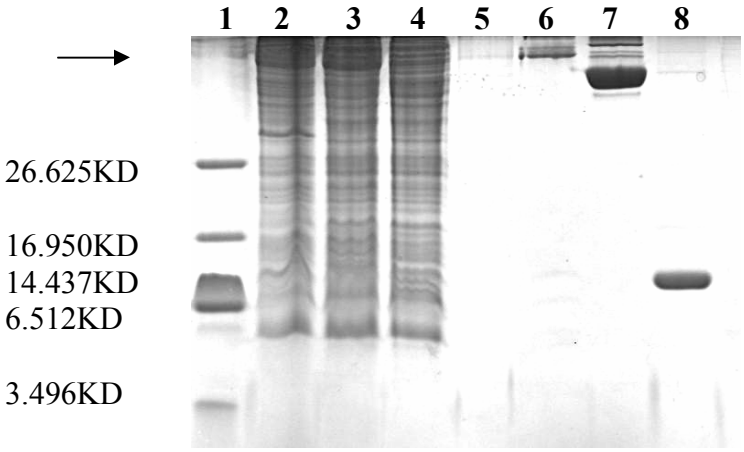


Figure3.6 SDS-PAGE of the fractions collected from each step in the Hfq expression and purification process. Lane1: Protein Marker (Bio-Rad 161-0326). Lane 2: Crude extract from cells. Lane3: Clarified crude extract from induced cells. Lane4: Chintin column flow through. Lane5: Chintin column wash. Lane6: SDS stripping of remaining proteins bound to chintin column before DTT induced cleavage. Lane7: SDS stripping of remaining proteins bound to chintin column after DTT induced cleavage. Lane8: eluded Hfq after DTT induced cleavage. The arrow indicates the fused protein of intein tag and Hfq.

The SDS-PAGE of the fractions at each step was shown in Figure 3.6. Lane2 and lane 3 are the crude cell extract before and after 20000g centrifugation respectively. After loading to the column, the band correspondent to the fused protein Chitin binding domain-Hfq (as indicated by arrow in Figure3.6) was greatly weakened, suggesting most of the protein was bound to the column. After DTT induced cleavage, the Chitin binding domain remains in the column (lane 7) and the elute fractions contain the target Hfq correspondent to ~11 KD (lane 8).

Centriprep-10 was used to concentrate the 60ml collected protein fraction to the final 2-3 ml. The excessive DTT in the protein sample was removed by washing with Column buffer (0.5 M NaCl, 20mM Tirs-HCl; pH 8, 1mM EDTA) using Centrion-30. Since Hfq is a high affinity RNA-binding protein, a certain amount of cellular RNA can be associated with the protein, that may interfere with the afterward RNA binding study. To remove that, DEAE anion exchange chromatography was used to further purify the protein (Carmichael and Weber 1975). A 3ml DEAE resin was suspended in 20 volumes of TE (pH7.6) containing 0.6M NaCl twice for equilibration, and then packed into a small column. 20ml of three different buffers, Buffer A (TE pH 7.6, 0.6M NaCl), Buffer B (TE pH7.6) and Buffer C (0.05 M Tris-HCl pH7.5, 1mM EDTA, 5%(w/v) glycerol, 0.05M NH_4Cl) respectively were used to wash the column subsequentially. The Hfq protein sample was load directly into the column and wash out by buffer C. The collected fractions was concentrated by Centricon-30 and stored in buffer containing 0.5 M NaCl, 20mM Tirs-HCl; pH 8, 1mM EDTA. All other mutant Hfq proteins follow the same purification procedure as wild type Hfq. SDS-PAG (Figure 3.7) showed a band which correspondent to 11kD monomer for all proteins. A faint band at 66kD was occasionally

observed. It has been attributed to incomplete denaturation of the hexamer prior to loading the gel lane. Silver staining showed no other bands.

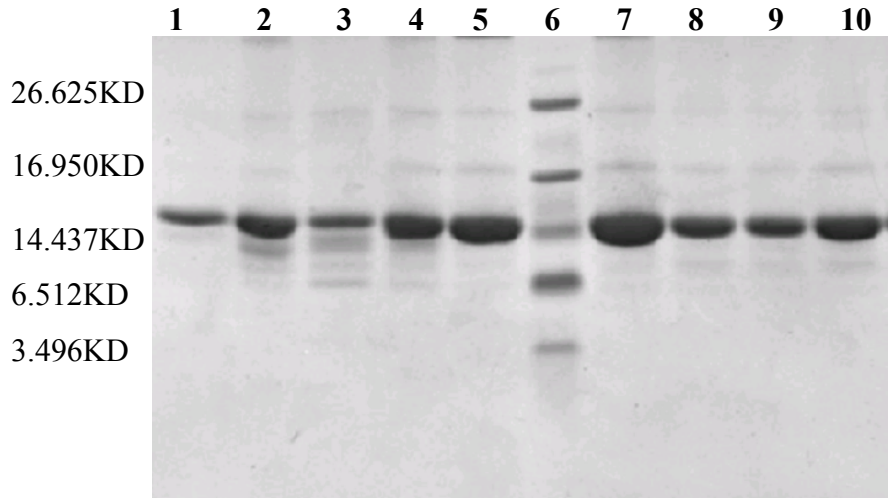


Figure3.7 SDS-PAGE of wild type and mutant Hfq proteins. Lane1: wild type Hfq, Lane2: F39A Hfq, Lane3: F42A Hfq, Lane4: F39A/F42A Hfq, Lane5: F39A/L12F Hfq, Lane6: Protein Marker (Bio-Rad 161-0326), Lane7: Q8A Hfq, Lane8: R16A Hfq, Lane9: Y25A Hfq, Lane10: K31AHfq.

UV absorbance at 276 nm of Hfq in 5.5 M Guanidium hydrochloride and aqueous solution were used to determine molar concentrations (Gill and von Hippel 1989). Extinction coefficients of 3850 M^{-1} cm^{-1} at 280 nm or 4250 M^{-1} cm^{-1} at 276 nm were employed for all Hfq's except Hfq-Y25A. For Hfq-Y25A the extinction coefficient employed was 2833 M^{-1} cm^{-1} at 276 nm. UV spectra were also used to assess RNA contamination. Pure Hfq was assumed to have a spectrum predicted empirically from the composition of its amino acids. The RNA contaminant was assumed to be polyA. Using experimental Hfq absorbance values at pairs of wavelengths from 260-280 nm this analysis gave 1 to 6 % RNA. Since the analysis of UV spectrum of ribonuclease A using these assumptions gave similar results, RNA contamination appeared small. CD spectra

from 200-300 nm were similar for the Hfq proteins and consistent with previous spectrum (Brescia et al 2003). To estimate the isoelectric point, PI and the protein charge at pH 8.0, the sequence of Hfq was entered into the Protein Calculator from http://www.scripps.edu/~cdputnam/protcalc.html. An isoelectric point 7.66 is estimated based on the pKa's of isolated residues. The charge at pH 8.0, the value for all buffers used to store Hfq, is estimated to be -0.4 based on the pKa's of isolated residues.

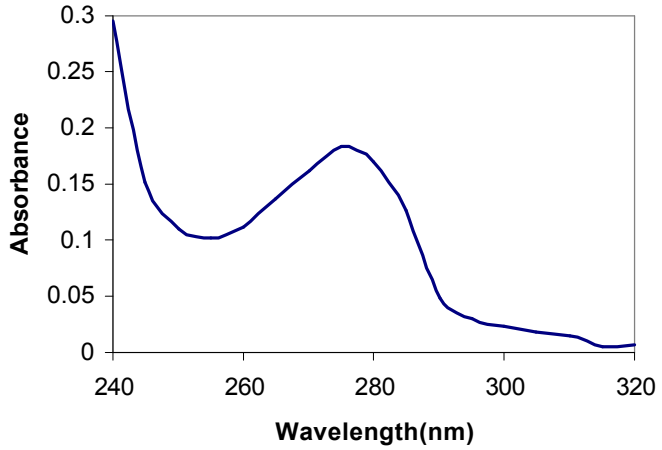


Figure3.8 UV spectrum of purified wild type Hfq protein

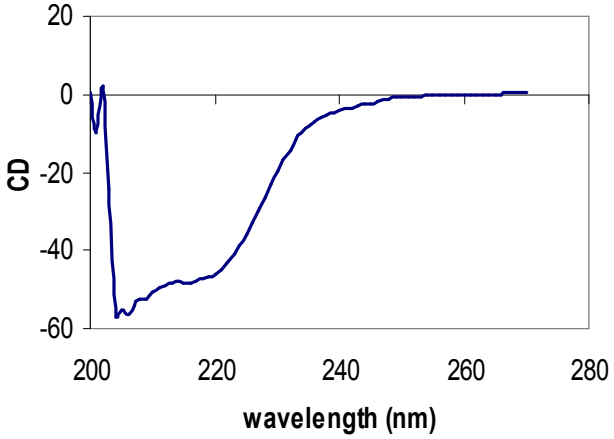


Figure 3.9 CD spectrum of purified wild type Hfq protein

HPLC Purification of the synthesized RNA oligomers

Three RNAs [A_{18} with 6-carboxyfluorescein (FAM) linked to the 5' end, and the 38 nt $DsrA_{DII}$ (5'-AACGAAUUUUUUAAGUGCUUCUUGCUUAAGCAAGUUUC-3'.) with Oregon Green-514 (OG-514) linked to its 5' end or FAM linked to its 3' end (shown in Figure 3.10)] were purchased commercially and purified by HPLC.

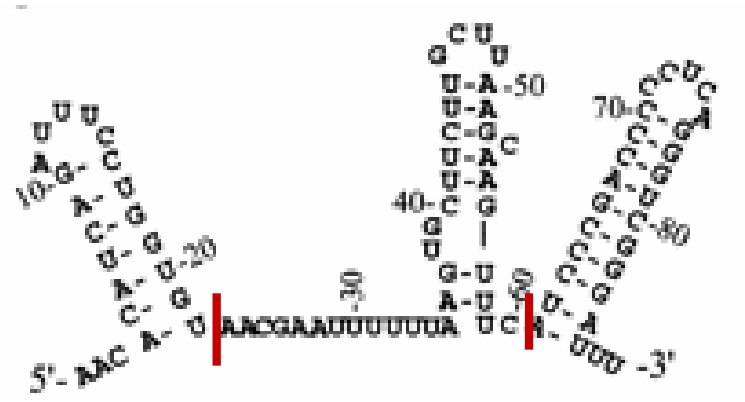


Figure 3.10 The predicted secondary structure of DsrA RNA
(the red bar indicates the region for second Domain)

HPLC Purification was done on an Agilent 1100 series machine with a manual injector, degasser, quaternary pump, and variable wavelength detector. The column used was a Dionex DNAPac PA-100 anion exchange column. All buffers were filtered before initial use with a 0.2 μm bottle-top filter. The system was set up such that buffer A was HPLC-grade water, buffer B was 0.25 M Tris-HCl, pH 8.0, and buffer C was 0.375 M sodium perchlorate. Buffer D was a storage buffer consisting of 80% water and 20% methanol. Prior to beginning the purification, each buffer was purged by pumping at 5 ml/min with the pump valve open to allow for the removal of any dissolved gases in the buffers. Water (100% buffer A) was then run through the column at a rate of 1.5 ml/min

with the pump valve closed for 5 minutes to ensure the removal of storage buffer. Finally, the conditions were set to 88% buffer A, 10% buffer B, and 2% buffer C for 10 minutes at a rate of 1.5 ml/min to prepare the column for injection.

Table 3.4 HPLC program for separation of RNA

Time (min)	Flow rate (mL/min)	BufferA (%)	BufferB (%)	Buffer C (%)
0.0	1.5	88	10	2
20.0	1.5	52	10	38
21.5	1.5	0	0	100
25.5	1.5	0	0	100
25.6	1.5	88	10	2

RNAs were deprotected according to the manufacturer's protocol prior to purification. In short, each tube was centrifuged briefly and 400µl of the provided 2'-deprotection buffer (100mM acetic acid, adjusted to pH3.8 with TEMED) added. Samples were pipetted up and down to completely dissolve the pellet and then vortexed for ten seconds, followed by centrifugation for ten seconds. Tubes were then placed in the Speedvac with the heat off and spun until dry. Following this centrifugation, the dry product was resuspended with DEPC-treated water and the tubes combined to give a final concentration of 150 to 250 OD_{260} / ml. Again, the volume of water in which the sample was resuspended was chosen to be a multiple of 100µl.

Injection was performed once the column had been prepared with the 88:10:2 mixture of buffers A, B, and C. The pump was set with the conditions shown in Table 2.1. The cycle time was set to 36 minutes to allow for column regeneration and reequilibration in preparation for additional injections. The detector was set at 260 nm and peaks were collected manually. Once all injections were completed, the column was

washed with 100% buffer A for 5 minutes to remove salt and then stored by running 100% buffer D through at 1.5 ml/min for 10 minutes.

Once purification was complete, desalting was necessary. The desalting method used a reverse phase desalting cartridge (Sep-Pak C-18, Catalog#020515, Waters, Inc.) It was necessary to equilibrate the cartridge prior to use. The plunger was removed from a 10 ml syringe and the SepPak cartridge attached. 10ml of acetonitrile was added to the syringe and the plunger replaced. The acetronitrile was passed through the cartridge at a rate of 2-3 drops per second. The cartridge was then removed, followed by the syringe plunger, and the cartridge returned to the syringer between each flush with solution. 10ml of a 50% acetonitrile: 50% 0.3M sodium acetate, pH 5.2 solution was added to the syringe, the plunger replaced, and the liquid passed through at the same rate as before. Next, 10 ml of 0.3 M sodium acetate was used to equilibrate the cartridge.

The volume of nucleic acid solution collected from the HPLC was estimated and 0.1 volumes of 3M sodium acetate, pH 5.2 added to the tube. This was vortexed to mix well, then added to the syringe used for desalting. This solution was passed through at a rate of 1 to 2 drops per second to allow time for the nucleic acid to bind to the C18 packing, and recollected. To ensure that as little nucleic acid as possible wound be lost, the solution was again added to the syringe and run over the SepPak cartridge. The cartridge was then washed with 5 ml of 50 mM triethylammonium bicarbonate (TEAB) at a rate of 2 to 3 drops per second. Finally the RNA was eluted with 5 ml of a solution of 35% methanol, 35% acetonitrile, and 30% 50 mM TEAB. This solution was collected into four approximately equal fractions and then spun to dryness in the SpeedVac with

the heat off. RNA was then dissolved into DEPC made TE buffer. Their concentrations were determined by absorbance readings at 260 nm and 494nm.

Sequence analysis of Hfq

BLAST searches were carried out against the Microbial Genome database at NCBI (http://www.ncbi.nlm.nih.gov/cig-bin/Entrez/genome_table_cgi), using *E. coli* Hfq or *A. fulgidus* Sm-1 sequences as the query. The multiple alignments were constructed by the CLUSTALW (http://www.ebi.ac.uk/clustalw/) program using the output of BLAST searches.

Gel mobility shift assay

For some gel shift assay, 10µl FAM-A_{18} or a fluorescent-labeled $DsrA_{DII}$ was mixed with the indicated amounts of wild type and mutant Hfqs respectively to give a total volume of 20µl in a binding buffer of 20mM Tris(pH 8.3), 0.5 M NaCl. The $DsrA_{DII}$ was heated for 3 min at 85 ^{o}C and quick cooled on ice for 10 min prior to mixing with the proteins. The reactions were incubated at 37 °C for 10 min, whereupon 2µl loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 40% glycerol) was added. Samples were run on a 6% PAG in 0.5X TBE with 4% glycerol at room temperature at 120 volts. After electrophoresis, the gel was scanned using a Fuji-film Image Reader FLA-3000 using the excitation and emission modes of 473nm and 520nm respectively. Band intensities were evaluated using AlphaImager^{TM} 950 software. Gel mobility shift experiments were also carried out with ^{32}P 5'-end labeled $DsrA_{DII.}$ The $DsrA_{DII}$ molecule was 5'-end labeled using ^{32}P-ATP and T4 polynucleotide kinase (New England Biolab)

and purified from the ATP using a Bio-Spin P-30 column (Bio-Rad Inc). Concentration of the purified RNA was determined by comparing its band intensity with the labeled unpurified RNA of known concentration.

Table 3.5 Reaction for radio-labeling $DsrA_{DII}$

$DsrA_{DII}$ (5μM)	2 μl
10 X buffer	10 μl
H_2O	74 μl
ATP	3 μl
T4 polynucleotide kinase	1 μl
Total	100 μl

Fluorescence quenching study

Fluorescence measurements were performed on an ISS spectrofluorimeter. The binding buffer used for all measurements contained 20mM Tris (pH 8.3) and 0.5 M NaCl. Emission spectra of Hfq's intrinsic fluorescence due to its tyrosines were scanned from 290-390 nm with excitation at 277 nm. Fluorescence quenching experiments were carried out with 50μl of wild type or mutant Hfq at 5 μM (in hexamer) and serially adding 2μl aliquots of 10μM (unlabeled) A_{18} until the saturation of quenching was reached. The observed intensities were corrected for loss of signal due to dilution. If F^0 is the corrected fluorescence of the free protein and F is the signal when an amount of A_{18} is added, the percent quenching, Q, is expressed as $Q = (F^0 - F)/ F^0$. The possibility of the inner filter effect disturbing the measurement was examined by adding aliquots of A_{18} to a tyrosine solution with similar fluorescence intensity as the Hfq solution. No change in the tyrosine emission intensity was observed. Titration of Hfq with a DNA oligomer also gave no significant quenching (<5%).

Fluorescence anisotropy measurements

Fluorescent molecules preferentially absorb light in a particular plane, termed the absorption dipole. The maximum absorption occurs when the absorption dipole moment of a chromophore is parallel to the plane of polarized light. Likewise, the emission dipole also has an intrinsic geometry, resulting in polarization of fluorescence emission. However, because the excitation and emission dipoles are not parallel, emitted light is partially depolarized relative to the excitation plane. This is referred to as intrinsic depolarization. Additional depolarization (extrinsic depolarization) that occurs during the lifetime of the excited state may result from rotation of the molecules and fluorescence energy transfer. It is the measurement of extrinsic polarization that provides the basis for quantifying molecular complex formation. For example, if a molecule rapidly and randomly reorients itself in solution before emission, the emitted light becomes additionally depolarized. Motion leading to depolarization in this case is rotational, because translational motion will not decrease polarization of fluorescence emission. In general, the rotational motion of small molecules in solution is greater than that of large molecules, due to the increased overall mobility of the smaller species (Figure 3.11)

Fluorescence anisotropy measurements were carried out at 23 ^{o}C using 20mM Tris (pH8.3) and 0.5 M NaCl as the solvent. The L-format was employed with the excitation monochromator at 490nm and emission monochromator at 517 nm. Anisotropy values were obtained from the average of 10 iterations using an integration time of 10s for each measurement. 2 mm slits were employed (16 mm bandwidth). Wild type and mutant Hfqs was serially titrated into a 3 ml cuvette that contained 2 ml of 2 nM of FAM-A_{18}, or 5 nM of 5'-OG514 or 3'-FAM labeled $DsrA_{DII}$. The total fluorescent intensity and

emission spectra of A_{18} showed no significant change with Hfq binding after accounting for dilution. The two $DsrA_{DII}$ molecules showed a 5% decrease in intensity in addition to the dilution effect after adding 600nM Hfq.

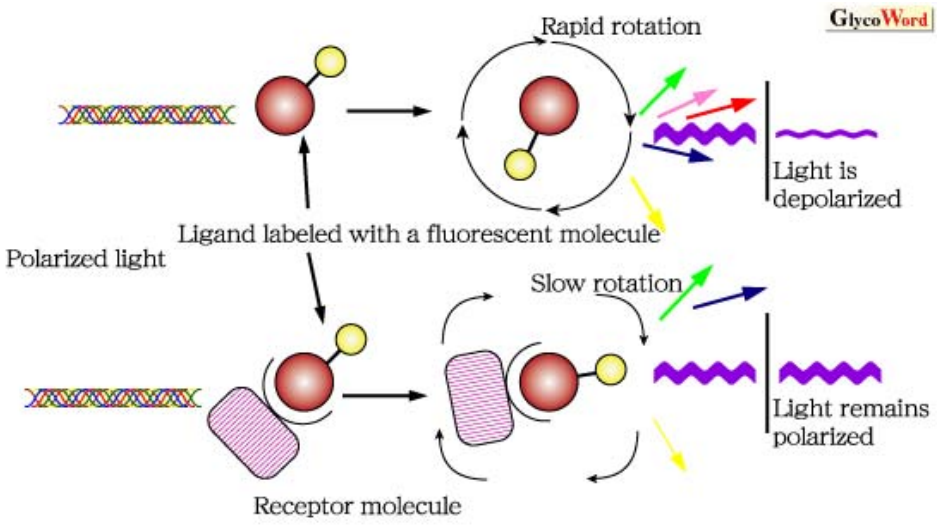


Figure 3.11 Schematic representation of the determination of receptor-ligand interaction by fluorescence anisotropy

Two models were employed in the analysis of the equilibrium binding of Hfq with the RNA molecules. One model assumed a one to one complex forms between a Hfq hexamer and RNA molecule. An equation describing the fluorescence anisotropy in terms of the dissociation constant K_d and other parameters of the experiment can be derived (Lundblad *et al.*, 1996) and is given by eq. (1).

$$A = A_f + (A_b - A_f)\ \{\alpha - [\alpha^2 - 4\ R_t\ P_t]^{1/2}\}/2\ R_t$$
$$\text{with}\ \ \alpha = R_t + P_t + K_d \qquad (1)$$

A is the measured anisotropy of the fluorescent RNA during the titration, A_f and A_b are the anisotropy of the free and bound RNA respectively, and R_t and P_t are the total

concentrations of RNA and Hfq hexamer respectively. A non-linear least squares fit of the equation to the data was made fitting the parameters A_b and K_d.

The second model assumed that Hfq hexamer binds RNA in a two-step reaction. The binding reaction is described by a dissociation constant K_1 for binding one Hfq hexamer to RNA, and a dissociation constant K_2 for binding a subsequent Hfq hexamer.

$$R + P \leftrightarrow RP \qquad \text{with} \quad K_1 = (R)(P)/(RP) \qquad (2a)$$

$$RP + P \leftrightarrow RP_2 \qquad \text{with} \quad K_2 = (RP)(P)/(RP_2) \qquad (2b)$$

Data was fit to the second model using the BIOEQS program (LeTilly and Royer, 1993). This numerical algorithm performs a least squares fit to the data fitting the free energies corresponding to K_1 and K_2, and the anisotropies of the free RNA, the 1:1 and 2:1 protein-RNA complexes. In general, the anisotropy of the free RNA was fixed to the experimental value, and the remaining four parameters fit to the data.

Results and Discussion

Sequence-Structure Analysis: **identifying potential RNA binding sites in Hfq**

Previous work has shown that RNA-binding proteins divide into two main classes based on their mode of RNA recognition (Draper, 1999). Groove binding proteins were designated as class I and position a secondary structure element such as a α-helix or a loop into a groove of an RNA helix. β-sheet binding proteins, designated as class II, utilize β–sheet surfaces to create binding pockets that bind unpaired RNA bases. Hfq is a β-sheet rich protein (a α-helix followed by five β strands) suggesting that its RNA

binding sites have characteristics of class II RNA-binding proteins. This form of binding was observed in the crystal structure of AU_5G oligomer with the *Staphylococcus aureus* Hfq (Schumacher *et al.*, 2002). Statistical analysis of 32 protein-RNA complexes also indicate that RNA-binding proteins have a preference for contacting guanine and uracil and that the residues lysine, tyrosine, phenylalanine, isoleucine and arginine have the highest propensities at RNA binding sites (Jones *et al*., 2001).

Using the above observations as guidelines, we sought to identify potential RNA binding sites of *E. coli* Hfq by determining surface locations containing residues with high propensities at RNA binding sites (i.e., Lys, Phe, Tyr, and Arg) and that are highly conserved among bacterial Hfqs. Inherent in this analysis are the assumptions that the functional form of Hfq is the hexamer, the structures of Hfq are very similar, and the RNA binding sites of different Hfqs are at the same locations. A multiple alignment of Hfq sequences was built based on a search of 354 completed or partially completed eubacterial genomes available at the NCBI database using BLAST (Altschul *et al.*, 1997). The core of the *E. coli* Hfq sequence (from residues 1 to 72) was used as the query sequence. More than half of the bacterial genomes contain at least one gene that is readily identified as Hfq. Figure 3.12 shows the alignment of Hfq sequences from amino acid positions 1 to 67 for a limited number of microbial genomes separated into bacterial groups by phylogenetic analysis (Sun and Wartell, 2002). In a few bacterial species two or three distinct copies of Hfq were found. They are designated by letters A, B, C.

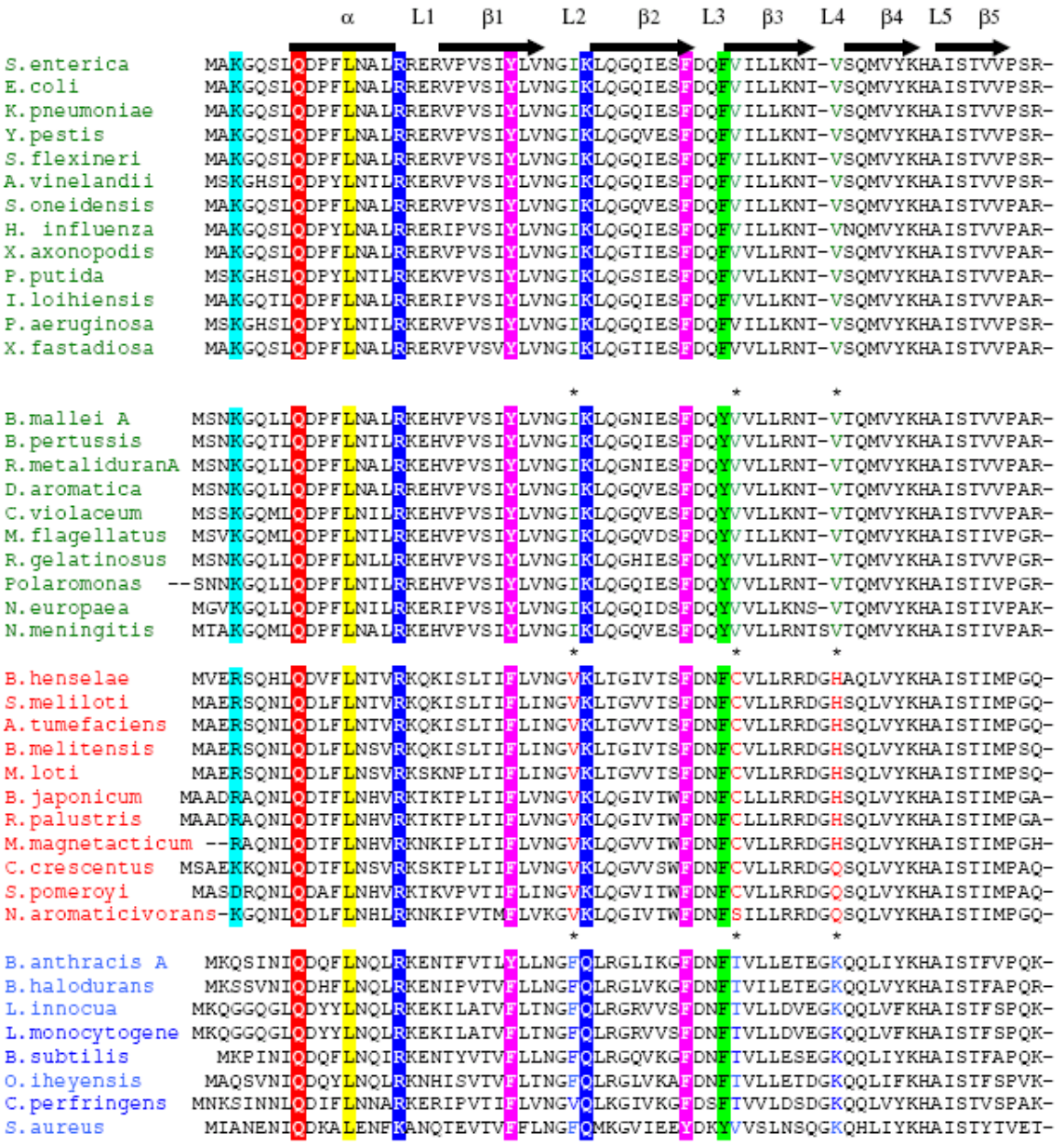


Figure3.12. Multiple alignment of Hfq conserved domain from representative bacterial genomes. The bar and arrows above the sequences represent the secondary structure elements α helix and β strand respectively. The numbering on the bottom is based on the *E. coli* Hfq sequence. Hfq sequences are subdivided into different bacterial groups accorded to the phylogenetic analysis. Green: γ and β-proteobacteria; Red: α-proteobacteria, Blue: Low-GC gram-positive; Three residues which serve as the fingerprint for each phylic group are marked by star and color-coded in the same way. The potential RNA binding residues which are mutated in this study are highlighted with the different color background. For full multiple alignment of the Hfq proteins see appendix I)

As has been previously noted (Moller *et al*., 2002; Zhang *et al.,* 2002) a number of positions have strikingly conserved residues or residues with similar chemical properties. Focusing on amino acids with high propensities at RNA binding sites, we note that Phe or Tyr at positions 11, 25, 39, 42, and 55, and Lys or Arg at 3, 16, 17, 31, 47 and 56 are highly conserved. Figure 3.13 shows the exposure of these residues on the proximal and distal surfaces of *E coli* Hfq as well as the edge-on or side view of the structure. Residues in only one of the six subunits are shown for clarity.

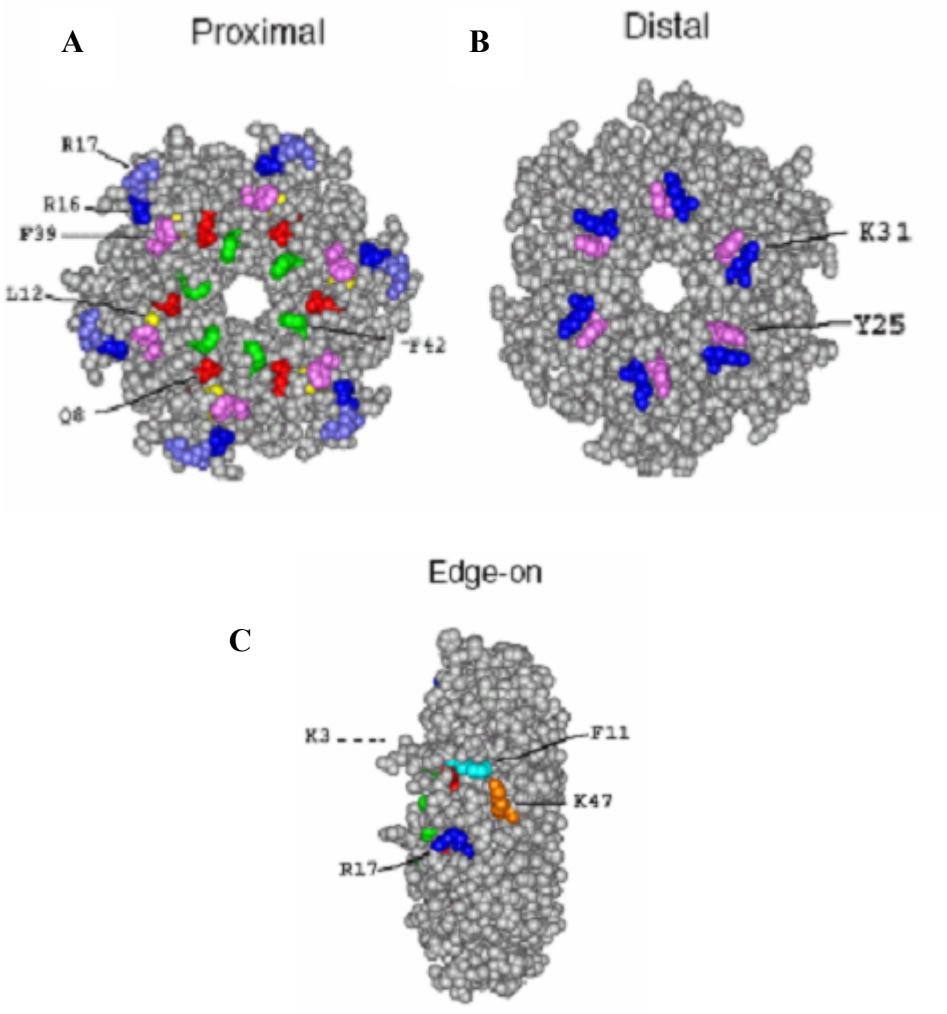


Figure 3.13 Space-filling representations of *E coli* Hfq.

Phe 11 is located near the outer rim of the structure. It is not visible when viewing the proximal or distal surfaces but is exposed in the side view. Lys 47 is also at the outer rim of the structure and its side chain atoms in subunit B are ~10 A from side chain atoms of Phe11 in subunit A (Figure 3.13C). This Phe11-Lys47 pair is reminiscent of the Phe and Lys pairs on the outer rim of TRAP protein subunits that interact with RNA bases (Antson *et al.,* 1999). Viewed from the proximal surface, Phe42, Tyr55 and Lys56 cluster near the cavity, and Phe39, Lys16 and Lys17 form a radially oriented patch near the outer part of this surface (Figure F.13A). Lys3, which could not be recognized in the crystal structure of *E. coli* Hfq (Sauter *et al.,* 2003), extends from the proximal surface. Tyr25 and Lys31 are adjacent to each other on the distal surface.

Evidence that several of these residues and surface sites are involved in Hfq binding to RNA is indicated from previous work. The crystal structure of *S. aureus* Hfq and AU_5G shows residues at or adjacent to positions corresponding to 42, 55 and 56 interacting with this RNA (Schumacher *et al.*, 2002). Mikulecky *et al*. showed that although *E. coli* Hfq binds much more weakly to AU_5G (Kd ~ 2 μM) (Mikulecky *et al.,* 2004) than *S. aureus* Hfq (Kd~ 50 nM) (Schumacher *et al.*, 2002), Y55A and K56A mutations of *E. coli* Hfq each reduce affinity to this RNA by about 10 fold. We note however that the mutation F42A in *E. coli* Hfq did not significantly affect Hfq binding to AU_5G or DsrA. Substitution of a number of residues for Lys3 disrupted Hfq function in Qβ RNA virus replication (Sonnleitner *et al.,* 2004). On the distal surface of *E. coli* Hfq, mutations at Tyr25 and nearby Ile30 reduced Hfq affinity to A_{27} implicating this surface in polyA binding.

Covariance of residue properties at spatially adjacent positions

It was noted during the alignment of Hfq sequences that approximately 4 % of the Hfq sequences did not have a Phe or Tyr at position 39. Figure3.14 lists six of them. In each case a smaller side chain replaced the aromatic residue expected at position 39 and simultaneously Tyr replaced the highly conserved Leu at position 12. These two positions, 12 and 39, are adjacent to each other in the Hfq structure at the same radial distance from the center of the hexamer (Figure3.13). A covariance of side chain properties appears to be needed at these highly conserved positions. This observation supports the notion that an aromatic residue at or near position 39 is important for Hfq function.

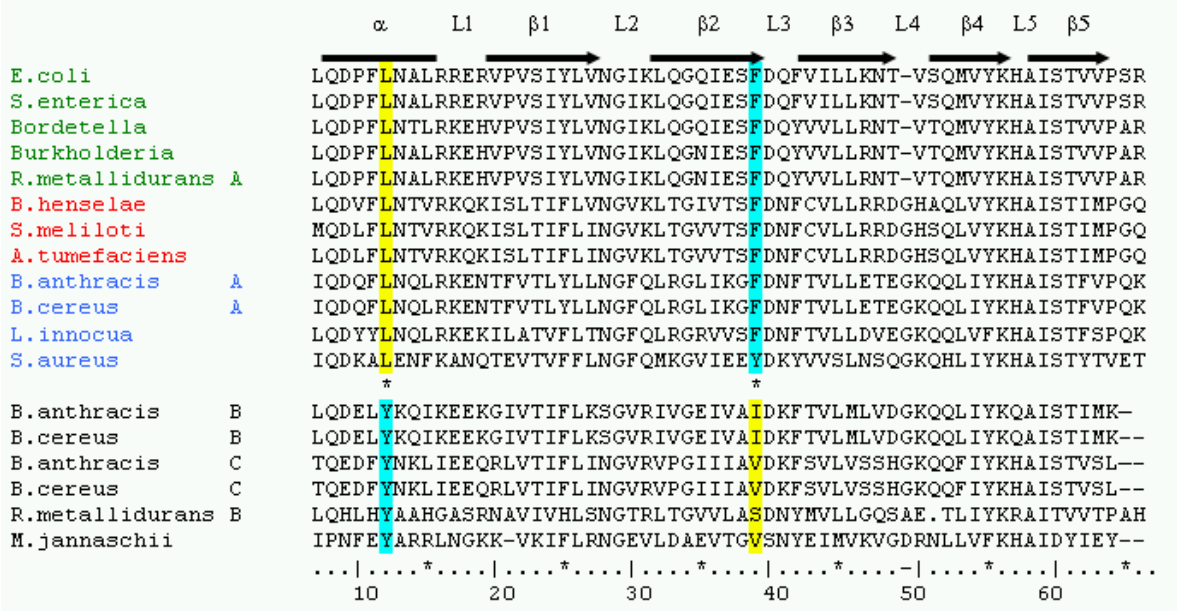


Figure3.14 The Hfq sequences which involve amino acid covariance in L12 and F39.

The covariance of an aromatic residue and small aliphatic residue at positions 39 and 12 observed in bacterial genomes was also observed in *M.jannaschii*, the first and only archeael specie found so far to contain a Hfq gene and no LSm gene (Valentin-Hansen *et al.*, 2004). A phylogenetic analysis shows this *M.jannaschii* gene belongs to a single phylic clade in the evolutionary tree of Hfq, separated from other bacterial species. Thus there is no evidence for lateral gene transfer from bacteria. The covariance exhibited by this archeal copy of Hfq as well as several bacterial Hfq genes suggests it is functionally important.

In addition, Results described in Appendix II indicate that a conserved aromatic group at the surface location of residue 39 in Hfq finds a close counterpart in structurally similar Lsm and Sm proteins and for some of the Lsm proteins, the amino acid covariance is also observed at the same locus on the surface of Lsm.

The effect of Hfq mutations on RNA binding

Guided by the above analysis we examined the effect of several Hfq surface residues on RNA-binding. Eight mutant *E coli* Hfq proteins were constructed and expressed as described in Material and Methods. Six of the mutant proteins, designated Hfq-Q8A, Hfq-R16A, Hfq-F39A, Hfq-F42A, Hfq-F39A/F42A and Hfq-L12F/F39A, are on the proximal face of the hexamer. Two mutant proteins, Hfq-Y25A and Hfq-K31A, are located in the distal face. The two RNA targets employed were $DsrA_{DII}$ (domain II of DsrA, nucleotides 23 to 60 (Brescia *et al.,* 2003) and A_{18}. $DsrA_{DII}$ is the major binding domain of a sRNA while A_{18} represents the 3' terminal segment of several mRNAs known to be regulated by Hfq (Folichon *et al.,* 2005; Le Derout *et al.,* 2004; Hajnsdorf E

et al., 2000). Domain II of DsrA contains the U-rich region implicated in Hfq-DsrA binding, but is missing RNA regions that increase affinity for wt Hfq. It was thought its interaction with Hfq might be more sensitive to specific mutations.

Gel Mobility Shift Assay of Hfq-RNA Binding

Figure 3.15 shows gel shift experiments of wt Hfq and several mutant Hfq's with A_{18}. The wt Hfq, Hfq-F39A and Hfq-F42A show similar behavior. Increasing amounts of protein initially shifts A_{18} to a low mobility complex (C_1) and then to the complex labeled C_2. Gel shift experiments of the other mutant Hfqs were similar to wt Hfq with two exceptions. Hfq-K31A (bottom panel, Figure 3.15) showed only a small decrease in intensity of the free A_{18} band with increasing Hfq_6 and no well-defined complex bands. A similar result was obtained the Hfq-Y25A. Residues Tyr25 and Lys31 are situated on the distal face of the hexamer structure (Figure 3.13). Tyr25 and the adjacent Ile30 have already been identified to be important for binding A_{27} (Mikulecky *et al.,* 2004). Our results confirm the importance of Tyr25 in binding oligoA, and demonstrates that Lys31 also interacts with this sequence. We note that *in vivo* studies have shown that substitution of Lys31 with Arg, Cys, Glu, Gly, Phe, Pro or Leu suppresses the functional role of *E coli* Hfq in phage Qβ replication, while substitution with His, Gln and Tyr allow function in this assay (Sonnleitner *et al.,* 2004). Examining Hfq sequences in other species shows Gln replacing Lys at this position in all low GC gram-positive species. Comparison of the two amino acids Gln and Glu indicates the charge does matter in RNA-binding at this position, but this property is not the only factor. Otherwise, one

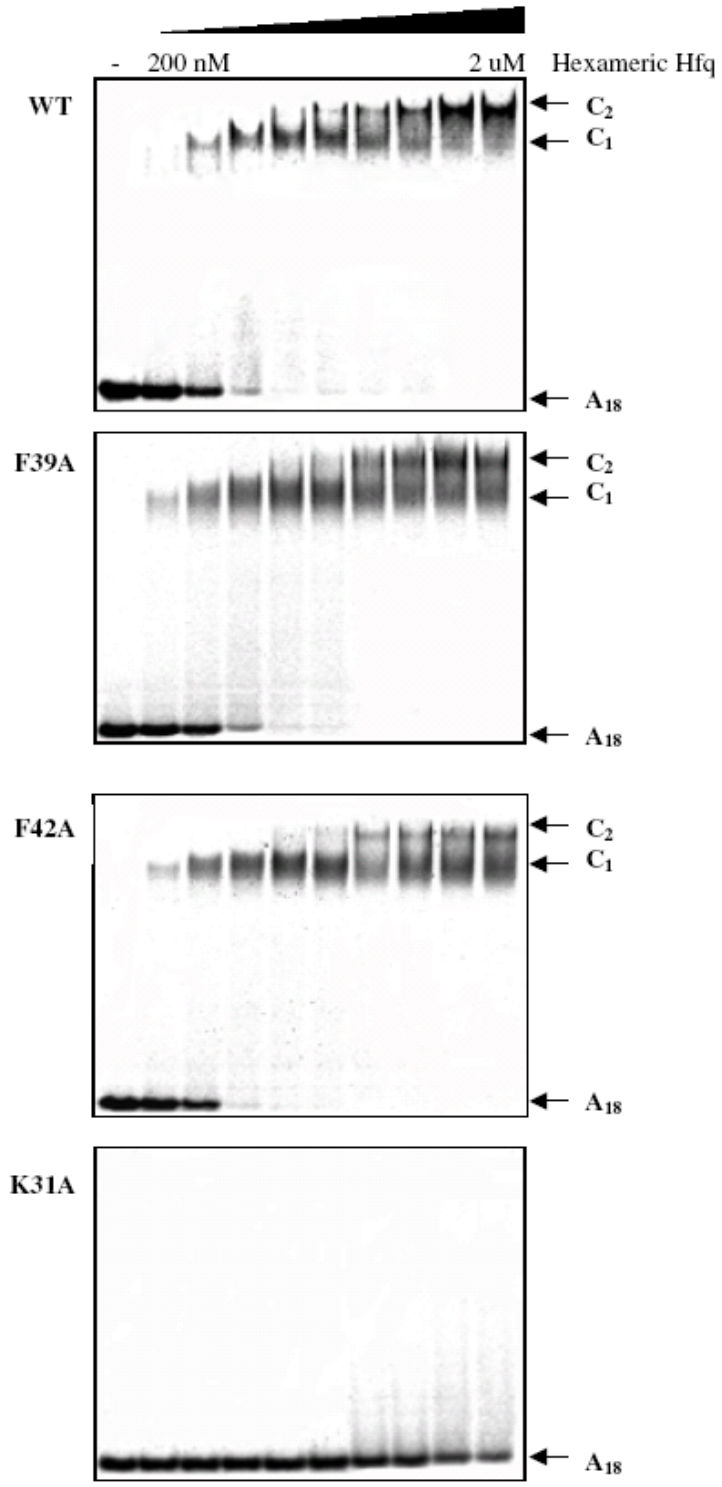


Figure 3.15 Gel shift experiments showing the binding of wild type and three mutant Hfq proteins to OligoA_{18} RNA. The concentration of Hfq_6 in each lane is 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 2.0μM respectively.

would expect Arg would allow the Hfq function as well. We note that two of the three functional substitutions are aromatic residues, namely His and Tyr, which make them capable of stacking with the base and stabilizing the interaction with RNA. Also these two residues are basic residues under certain conditions, so can form hydrogen bonds or/and electrostatic interactions with the RNA. The latter chemical property may explain why Phe does not function while Tyr does.

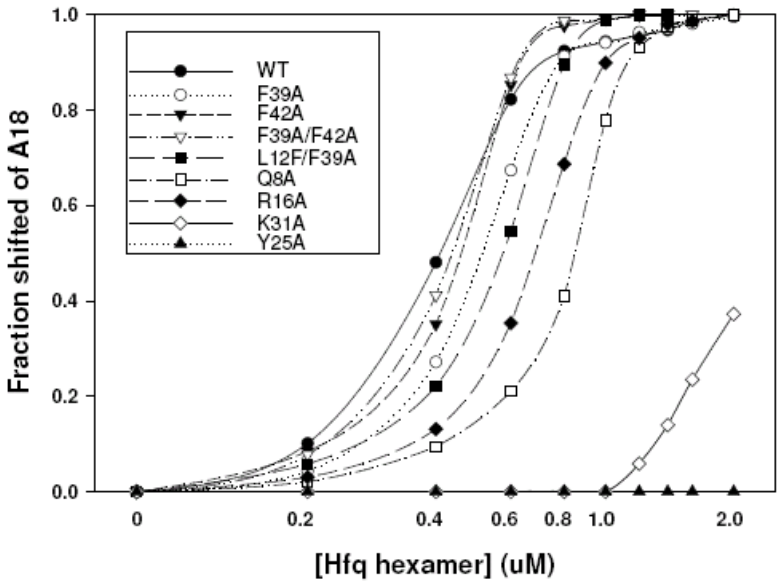


Figure 3.16 Quantification of the gel shift experiments of A_{18} for wild type and all the mutant Hfqs.

Figure 3.16 summarizes the gel shift binding results for A_{18}. The experiments conducted in Figure 3.15 employed the fluorescent FAM-A_{18} and used a total RNA concentration of 300 nM. Since this is well above the K_d of the wt Hfq-A_{18} complex measured from the gel shift assay and fluorescence anisotropy (results below) it should reflect stoichiometric binding. The titration of A_{18} with wt Hfq saturated the C_1 complex

at a ratio of Hfq_6 to A_{18} of approximately 2:1. Accurate evaluation of this ratio was difficult to establish since the C_1 band is not well defined, however two independent approaches described below are also consistent with a 2:1 ratio for Hfq hexamer to A_{18} in this complex.

Figure 3.17 shows gel mobility shift experiments of wild type and several mutant Hfq's to $DsrA_{DII}$. 2 nM of radioactively labeled $DsrA_{DII}$ was used in these experiments. wt Hfq initially shifts $DsrA_{DII}$ to a lower mobility complex (C_1), and then shifts the RNA to another complex (C_2) with increasing amount of Hfq. This phenomenon was previously observed in gel mobility shift assays using the complete DsrA (Mikulecky *et al*., 2004). The second panel of Figure 3.17 indicates that Hfq-F39A has a lower affinity for $DsrA_{DII}$ than wt Hfq. The intensity of the unbound $DsrA_{DII}$ band was observed at protein concentrations where no free $DsrA_{DII}$ was detected for wtHfq. Additionally the 'band' at the position of the complexes were more diffuse with intensity smeared below indicating that complexes of Hfq-F39A and $DsrA_{DII}$ dissociate more readily in the gel. The third panel shows results with Hfq-R16A which were similar to Hfq-F39A. The bottom panel shows the gel shift experiment of Hfq-L12F/F39A binding to $DsrA_{DII}$. The relative amount of the free $DsrA_{DII}$ band decreased more rapidly with additions of Hfq-L12F/F39A than with equal amounts of Hfq-F39A. This double mutation partially restored the binding defect caused by F39A

Figure 3.18 displays the fraction of bound $DsrA_{DII}$ as a function of Hfq concentration assessed from the change in intensity of free $DsrA_{DII}$ bands. The concentration at which $F_B = 0.5$ ($F_{0.5}$) for wt Hfq was 470 nM. Most mutant Hfq had a $F_{0.5}$ value between 550 nM and 1uM, however for Hfq-R16A the value was 2.6 uM,

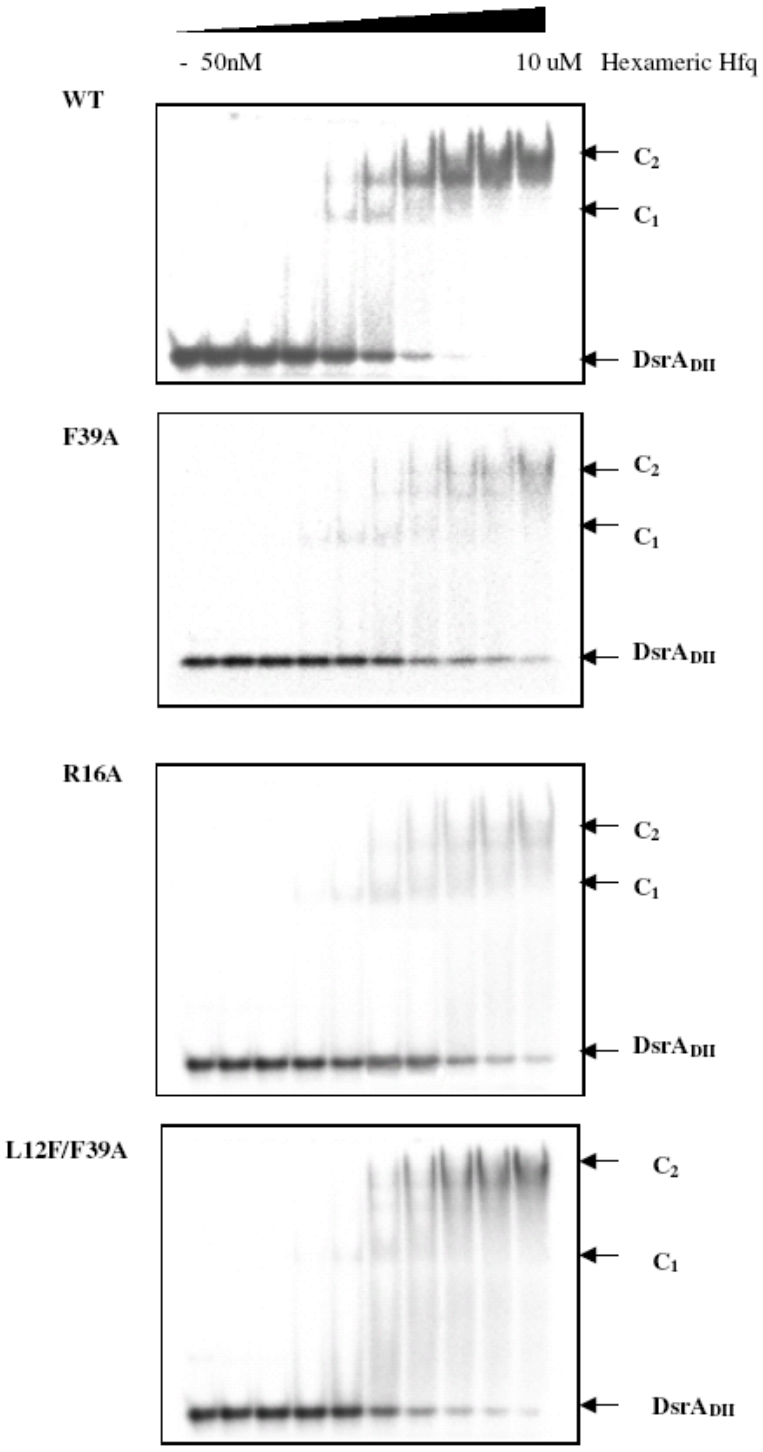


Figure 3.17 Gel shift experiments showing the binding of wild type and three mutant Hfq proteins to radio-labeled $DsrA_{DII}$ RNA. The concentration of Hfq_6 in each lane is 0.05, 0.1, 0.25, 0.5, 1.0, 2.0, 4.0, 6.0, 10.0μM respectively. The concentration of RNA is 2nM.

While the midpoint for Hfq-F39A and Hfq-F39A/F42A was 3.8 uM. The 6-8 fold increase in the apparent K_d for Hfq-R16A and Hfq-F39A compared to wt Hfq suggests that these sites interact with $DsrA_{DII}$. The observation that the double mutant Hfq-L12F/F39A increased the affinity for $DsrA_{DII}$ relative to Hfq-F39A supports the notion suggested from the bioinformatics analysis that an aromatic group at this location is involved in RNA binding.

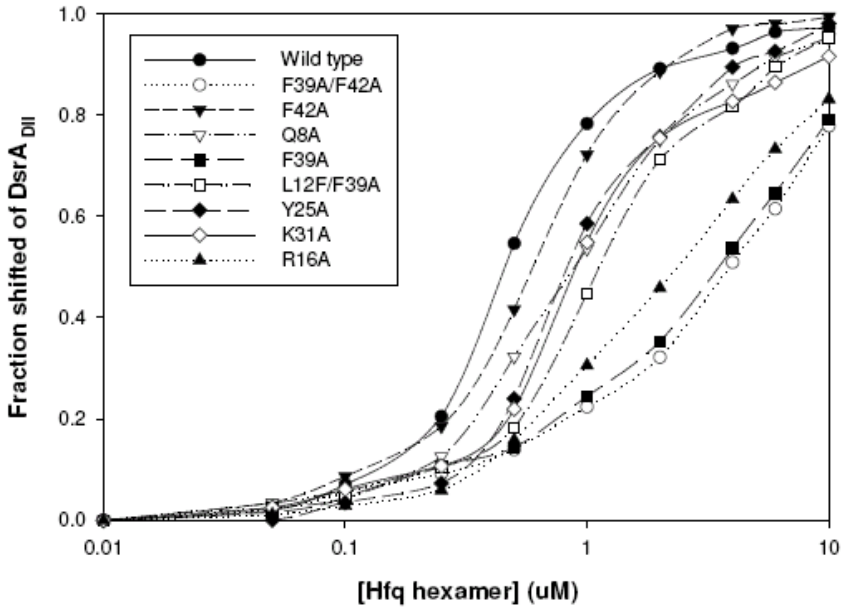


Figure 3.18 Quantification of the gel shift experiments of $DsrA_{DII}$ for wild type and all the mutant Hfqs.

It is worth noting that the effects of these Hfq mutations on $DsrA_{DII}$ binding are relatively small. No mutation showed the 100-1000 fold effect observed for the single site mutations Y25A and K31A on A_{18} binding. Gel shift experiments using fluorescent-labeled $DsrA_{DII}$ at 100 nM (Figure 3.19) produced better defined C_1 bands for some of the Hfq but the results were otherwise qualitatively similar to figure3.16.

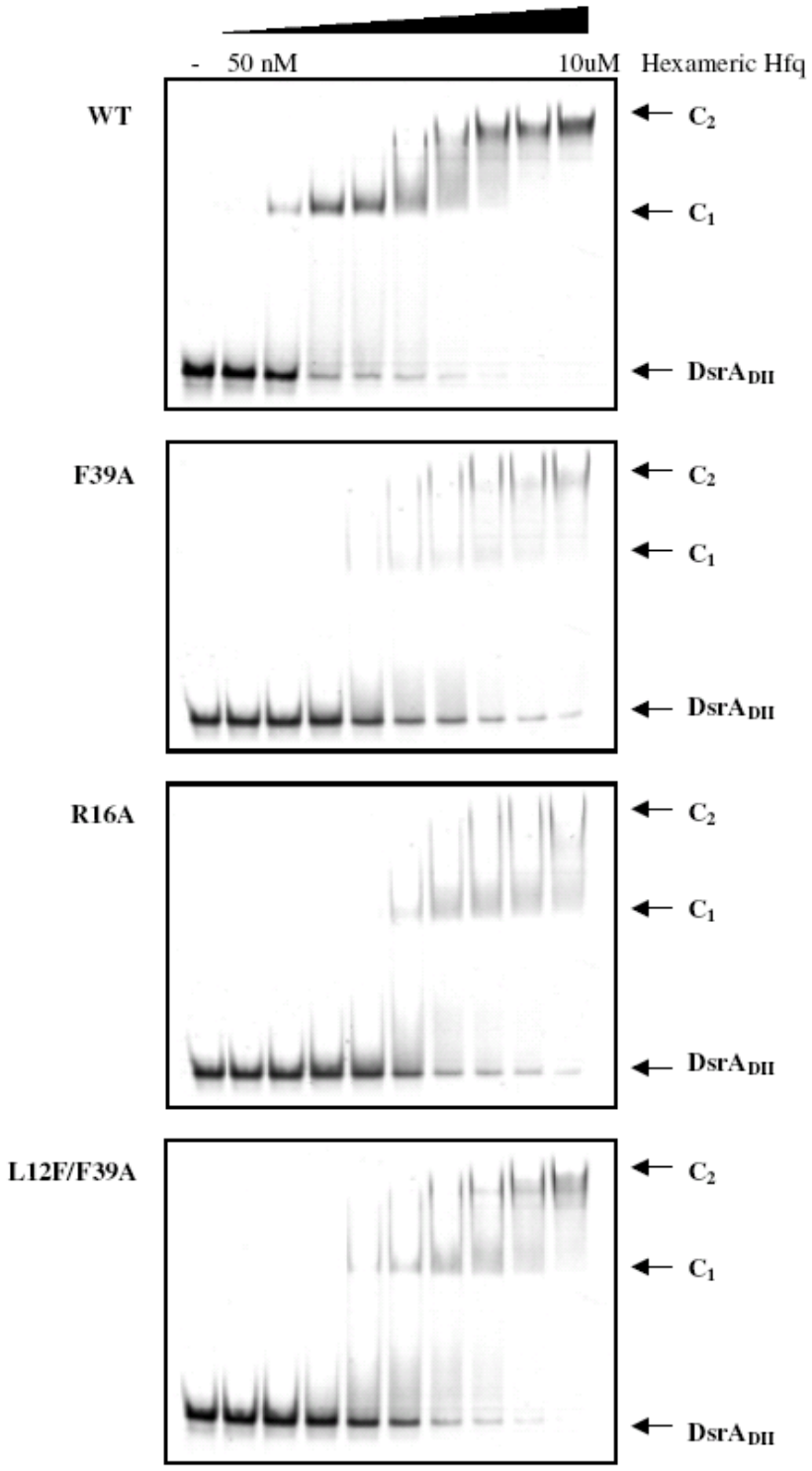


Figure 3.19 Gel shift experiments showing the binding of wild type and three mutant Hfq proteins to fluorescence labeled $DsrA_{DII}$ RNA. The concentration of Hfq_6 in each lane is 0.05, 0.1, 0.25, 0.5, 1.0, 2.0, 4.0, 6.0, 10.0μM respectively, fluorescence labeled DsrA is 100 nM.

Figure3.20 shows results from a gel shift assay carried out with 1 uM $DsrA_{DII}$ and wt Hfq. The C_1 complex reached a maximum intensity at a molar ratio of 2:1 Hfq_6 to $DsrA_{DII}$ and was constant until 3.5 uM Hfq_6 before its intensity decreased in conjunction with increased intensity of slower mobility complexes. A similar saturating ratio of 2:1 was observed previously (Lease and Woodson, 2004) for the strong binding complex formed by wt Hfq_6 and the complete DsrA at high concentrations.

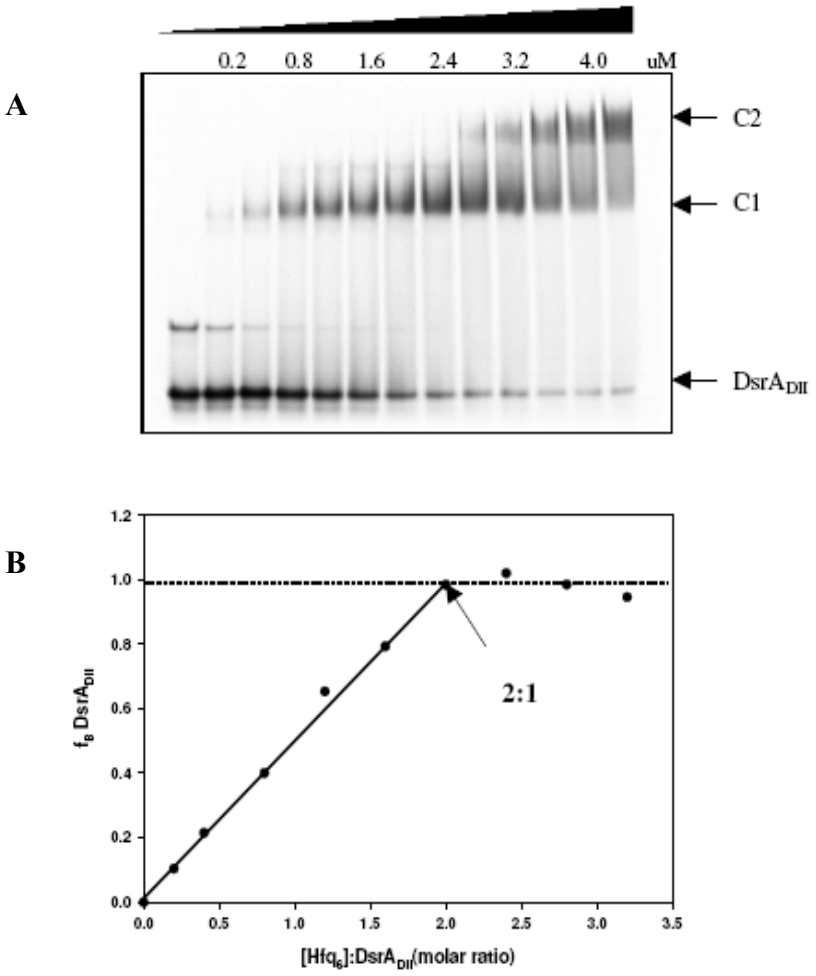


Figure 3.20 Stoichiometry of Hfq:$DsrA_{DII}$ complexes. $DsrA_{DII}$ RNA (1uM) was titrated with Hfq protein (A) and the fractional saturation of C1 complex was plotted against the ration of Hfq hexamer to DsrA (B)

Fluorescence anisotropy measurements of Hfq-A_{18} binding

Fluorescence anisotropy was employed to quantitatively analysis the binding affinity of the wt and mutant Hfqs with the two RNAs in solution (Lundblad *et al.,* 1996). The 5'-fluoresceinated A_{18} and Oregon green modified $DsrA_{DII}$ were serially titrated with wild type and mutant Hfq proteins in 0.5 M NaCl and 20 mM Tris-HCl (pH8.3). Figure 3.21A shows the anisotropy change for 2 nM A_{18} with increasing concentration of wt Hfq. The dashed line shows the non-linear least squares fit of a 1:1 model of Hfq binding to A_{18} to the data. The inability of this model to accurately fit the data indicates a more complex binding interaction.

A numerically based equilibrium binding algorithm, BIOEQS, was employed to analyze the anisotropy data using a model in which the RNA can bind one Hfq_6 and subsequently a second Hfq_6 (see Materials and Methods). The solid line in Figure 3.21A shows the least squares fit of this model to the data. The quality of the fit is considerably improved with a χ^2 of 1.3. The free energies evaluated indicate dissociation constants for binding the first and second Hfq_6 of 10.1 nM and 4.7 nM respectively. Confidence limit testing of the recovered free energies gave deep quadratic-well like plots of χ^2 vs free energy with minimums at both evaluated free energies.

Figure 3.21B shows the results of anisotropy experiments for two of the mutant Hfqs; Hfq-F39A and Hfq-K31A. As expected from the gel shift experiments, Hfq-F39A binds A_{18} similarly to wt Hfq, while Hfq-K31A shows very weak binding. Table 3.5 summarizes dissociation constants evaluated from the analyses of the anisotropy

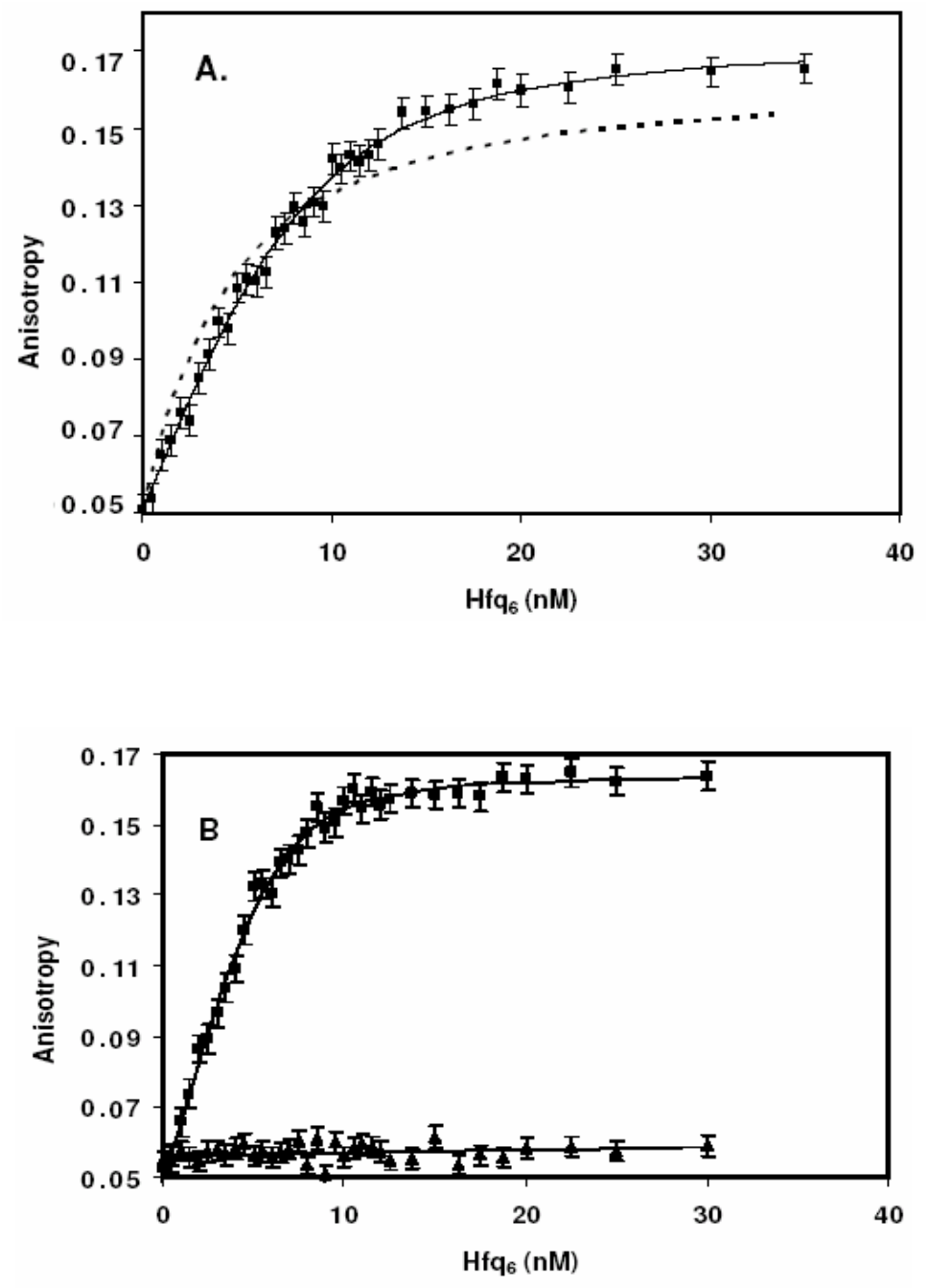


Figure 3.21 A. Fluoresence anisotropy data of wt Hfq binding to A18 with best fits for models of binding 1:1 complex dashed line, and model of 1:1 and 2:1 Hfq to A18 using BIOEQS. $delG_1$=10.9 kcal/mol, $delG_2$=22.25 kcal/mol **B**. F.A. data for binding of F39A Hfq to A18 (top) and K31A Hfq (bottom). Fits shown are for 2 component model for F39A giving delG1=10.98 kcal/mol delG2= 23.4 kcal/mol. For K31A the fit shows K31A barely binds the A18 since it's at baseline.

measurements. In all cases except for Hfq-K31A and Hfq-Y25A, the two step binding model was needed to produce a good fit to the anisotropy curves. For wtHFq and four of the mutant Hfq, the analysis yields binding constants for the second bound Hfq_6 that are 2-10 fold lower than K_1. Hfq-L12F/F39A, Hfq-Q8A and Hfq-F39A/F42A produced similar values for K_2 and K_1. For the two proteins K31A and Y25A, the anisotropy change was slightly higher than background. A rough estimate of K_d based on a 1:1 binding model gave K_d >1.3 uM for these two single site mutations indicating a ~1000X decrease in affinity.

Table 3.6 Equilibrium dissociation constants of wild type and mutant Hfq binding to A_{18}

Hfq	K_1 (nM)	K_2 (nM)
Wild type	10.1 ± 1.2	4.73 ± 0.8
F39A	8.7 ± 0.8	0.75 ± 0.7
L12F/F39A	4.4 ± 0.8	3.61 ± 0.8
F42A	3.1 ± 0.8	0.62 ± 0.7
F39A/F42A	8.3 ± 0.9	8.27 ± 0.9
Q8A	12.0 ± 1.2	14.7 ± 1.0
R16A	14.9 ± 1.3	3.0 ± 0.7
K31A	>1.6 uM	
Y25A	>1.3 uM	

Figure 3.22 shows the results of fluorescence anisotropy experiments of wt Hfq binding to $DsrA_{DII}$. The fluorescence anisotropy reached a plateau between ~600 nM and 1.1 uM Hfq_6 and then increased slightly to another plateau with additions of Hfq_6 above 1.1 uM. The overall anisotropy change was considerably less than observed for Hfq

binding to A_{18}. The second plateau although quite small was reproduced several experiments. Given the signal to noise of the data we could not discriminate we could not discriminate between 1:1 and 2:1 binding models for Hfq_6 binding $DsrA_{DII}$ based on the quality of fit. If one assumes a 1:1 binding model that least squares fit to the first plateau region using eq.[1] yield K_d = 470 nM (solid line, Figure3.22). If one uses all of the data (dash line Figure 3.22) the best fit K_d is approximately 1uM. Fitting either all of the data or just the first plateau region to the two step binding model using BIOEQS yielded a dissociation constant of $K_1 \approx 440$ nM for binding the first Hfq_6 and $K_2 \approx 1.9$ uM for binding a second Hfq_6. However confidence limit testing indicated K_2 could not be accurately resolved.

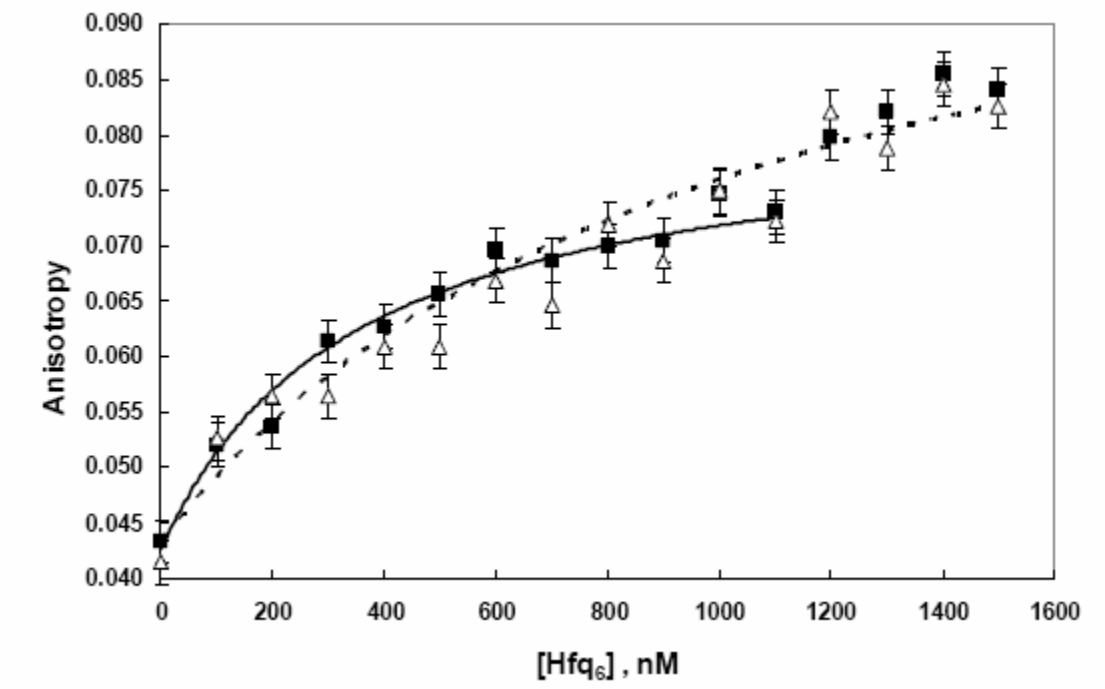


Figure 3.22 Fluorescence anisotropy of $DsrA_{DII}$ as a function of wt Hfq (■), and F39A-Hfq (Δ). Solid line is best fit of 1:1 model to wt Hfq data up to 1100 nM Hfq_6 and dashed line is best fit to 1500 nM.

Surprisingly, titration of $DsrA_{DII}$ with the mutant Hfq's produced anisotropy curves similar to that for wt Hfq. Anisotropy change as a function Hfq-F39A concentration is shown in figure 3.22 as open triangles. Data for the other mutant Hfq were similar. The data was not consistent with the eight fold reduction in affinity noted in the gel shift experiments. The discrepancy between the anisotropy measurements and gel shift experiments with $DsrA_{DII}$ is discussed below.

Fluorescence quenching measurements

A different approach that can be employed to investigate RNA binding to Hfq is to monitor the fluorescence quenching of Hfq as a function of RNA concentration. Previous work (DeHaseth and Uhlenbeck, 1980a, 1980b) showed that polyA binding quenches Hfq fluorescence. *E. coli* Hfq has a moderate intrinsic fluorescence due to its three tyrosines Tyr 25, Tyr55 and Tyr83. Tyr 25 is located on the distal surface of the hexamer. the apparent binding surface of polyA sequences. Tyr 55 is located near the central cavity close to the proximal surface, and Tyr 83 is in the unstructured C-terminal end.

Figure 3.23A shows the fluorescence emission spectrum of wt Hfq at 5 uM in hexamer. Titration of A_{18} decreased the fluorescence of Hfq until a saturation level (~73% at 305 nm) was reached. Quenching may result from the stacking of an adenine base with exposed Tyr and/or the transfer of the Tyr hydroxyl group proton to an acceptor on A_{18} and formation of the non-fluorescent tyrosinate anion. Figure 3.23B plots the percent quenching as a function of the ratio of A_{18} to Hfq. The value of $[A_{18}]/[Hfq_6]$ at the break point in the curve is close to 0.5, consistent with a 2:1 stoichoimetry of Hfq_6

to A_{18}. Similar results were obtained for the other mutant Hfq's except for Hfq-Y25A, Hfq-K31A and Hfq-Q8A.

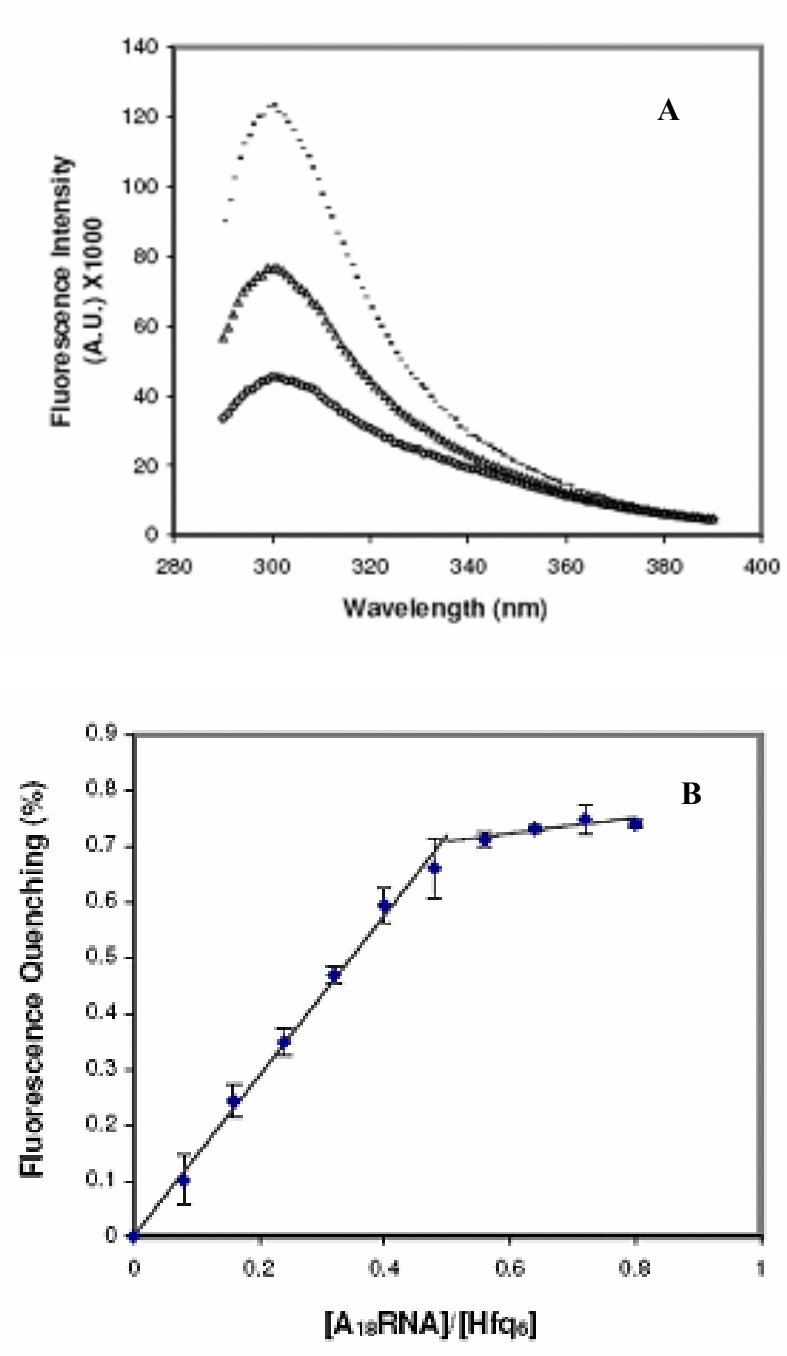


Figure 3.23 A Fluorescence Emission spectra of 50 µl wild type Hfq (30 µM monomer) with 0 µl (solid), 6µl (dot) and 10 µl (dash) oligoA18 RNA (10 µM). **B** Fluorescence titration of wild type Hfq with oligoA18 RNA.

Table 3.7 The saturated fluorescence quenching and stoichiometry of A_{18} for Hfqs

Hfq Proteins	Fluorescence Quenching (%)	$[A_{18}RNA]/[Hfq_6]$ Ratio
Wild Type	73 ± 2	0.50
F39A	73 ± 2	0.50
F42A	70 ± 1	0.50
F39A/F42A	71 ± 2	0.50
L12F/F39A	71 ± 1	0.50
Q8A	63 ± 2	-
R16A	72 ± 2	0.50
Y25A	13 ± 3	-
K31A	15 ± 3	-

The fluorescence of Hfq-Y25A and Hfq-K31A were quenched by only 13% and 15% respectively at saturating amounts of A_{18} and showed no clear break points in the titrations. These results indicated weaker binding of A_{18} to these mutated proteins, consistent with the data from gel-shift assay and anisotropy measurements. The fluorescence intensity of free Hfq-Y25A was about one-third that of wt Hfq at the same concentration. Reduce intensity is expected for this Hfq missing the surface exposed Tyr25. Free Hfq-Q8A had about 140% higher fluorescence intensity compared to wt Hfq at the same concentration. The latter observation can be understood in terms of two crystal structures of Hfq and the effect a Q8A mutation may have on Tyr 55 in a neighboring subunit(Schumacher *et al.*, 2002). Additionally a recent high-resolution structure of Hfq from *P. aeruginosa* shows a conserved hydrogen bond between the OH group of Tyr55 and OE1 of Gln8 that stabilizes the α-helix in Hfq (Nikulin *et al*., 2005). In wt Hfq, it is likely that the fluorescence of Tyr55 is quenched due to the hydrogen bond between its hydroxyl group and Gln8. The Gln to Ala substitution releases the

hydroxyl group of Tyr55, increasing the fluorescence originating from this residue and the overall fluorescence of the protein. The percent quenching of Hfq-Q8A at saturating amount of A_{18} was ~60% compared with compared to 73-75% for wt Hfq. This is qualitatively consistent with the notion that the quenching effect of A_{18} is due to its interaction with Tyr25. Since this residue contributes less to the overall fluorescence of Hfq-Q8A than to wt Hfq, the saturation quenching percentage is reduced. Additions of $DsrA_{DII,}$ rGU_5A and rU_{20} had a negligible effect on Hfq fluorescence.

Summary

The studies of A_{18} interacting with wild type and mutant Hfq proteins indicates that this RNA sequence interacts specifically with residues on the distal surface of the *E. coli* Hfq. The gel shift assay, fluorescence anisotropy, and fluorescence quenching experiments all showed that mutations on the proximal surface of Hfq (F39A, F42A, Q8A, and R16A, F39A/F42A, L12F/F39A) had relatively little effect on binding A_{18}, while single site mutations Y25A and K31A on the distal surface reduced affinity by 100-500 fold. Our results are consistent with the results obtained by Mikulecky et al (Mikulecky et al. 2004) indicating that A_{27} binding requires specific residues on the distal surface. We confirm their conclusion regarding the importance of Tyr25 for binding a polyA molecule, and demonstrate that a mutation to Lys31, which is adjacent to Tyr25, also results in a major reduction in Hfq binding.

The gel shift experiments of A_{18} in the presence of increasing Hfq concentration showed the existence of two protein-RNA bands we designated C_1 and C_2 (figure 3.14).

Using concentrations well above the K_d estimated from gel shift analysis (Mikulecky *et al*., 2004) and fluorescence anisotropy (this work) the C_1 band saturated at a stoichiometry of approximately 2:1 for wt Hfq_6 to A_{18} and then decreased as the C_2 band increased. Analysis of the fluorescence anisotropy data on Hfq binding to 2 nM A_{18} was consistent with a model in which two Hfq_6 molecules bind one A_{18}, each with nM affinity. In addition fluorescence quenching of wt Hfq by A_{18} at uM concentrations also indicates a 2:1 Hfq_6 to A_{18} stoichiometry.

All three of the above experimental approaches are consistent with a model in which two Hfq_6 bind to one A_{18} with strong affinity. Based on stoichiometry, the C_1 band in the gel appears to correspond to this complex. The absence of any significant shifted band with a faster mobility (corresponding to a 1:1 Hfq_6-A_{18} complex) implies that two Hfq_6 bind cooperatively to A_{18} in the gel. Analysis of the anisotropy data (Table 3.5) also suggests the second bound Hfq_6 has a greater affinity than the first for half of the Hfq's examined. We note that the sedimentation velocity centrifugation indicates that Hfq exists as a hexamer in the solution employed (Materials & Methods).

The relative mobility of the C_2 band in figure 3.14 and its appearance at the expense of the C_1 band indicates the existence of higher stoichiometry complexes of Hfq_6 and A_{18}. This phenomena, Hfq-RNA complexes with stoichiometry greater than 2:1 Hfq_6-RNA has also been observed in gel shift experiments with DsrA and RpoS (Lease and Woodson 2004). If the C_2 band is not an artifact of gel conditions one would expect it to exist at high ratios of Hfq_6 to A_{18} in solution. We note that the fluorescence quenching of Hfq's Tyr25 as a function of A_{18} additions produced a single linear slope until the 2 to 1 saturation point was reached (figure 3.22B). If a complex with higher Hfq_6 to A_{18}

stoichiometry occurs during the initial stages of this titration it was not evident from A_{18} interacting with the Hfq_6 distal surface.

The gel mobility shift studies of Hfq binding to $DsrA_{DII}$ indicated that the F39A and R16A mutations affect the affinity of Hfq to $DsrA_{DII}$ and imply that these residues interact with this RNA. These two mutations reduced the apparent K_d by 6 to 8 fold (assuming a 1:1 model) while other single site mutations altered the apparent K_d by 2-fold or less. The ability of the L12F/F39A-Hfq to partially restore the affinity of Hfq to $DsrA_{DII}$ supports the notion that Phe39 is involved in $DsrA_{DII}$ binding. It is also consistent with the bioinformatics analysis indicating that an aromatic residue at this surface location is important for Hfq function.

Using high concentrations of $DsrA_{DII}$ in the gel shift experiment resulted in better defined bands corresponding to RNA-Hfq complexes. As observed with A_{18}, two such bands were observed. The 2:1 stoichiometry determined for the stronger binding Hfq_6-$DsrA_{DII}$ complex is the same as observed with A_{18} as well as in previous studies on the 87 nt DsrA and a 140 nt RpoS RNA (Lease and Woodson 2004). This common value for several RNAs differing in length and sequence suggests that binding of *E. coli* Hfq_6 to RNA induces the binding of a second Hfq_6.

In contrary with the gel shift studies, the fluorescence anisotropy experiments did not show a significance difference in the binding of wt and mutant Hfqs to $DsrA_{DII}$. Although this discrepancy is not currently understood, several factors provide possible explanations for this difference. The anisotropy experiments have the advantage of being an equilibrium measurement in solution however the anisotropy change observed for the fluorescently labeled $DsrA_{DII}$ upon Hfq binding was relatively small. Small differences in

affinity between wt and mutant Hfq for $DsrA_{DII}$ may go undetected. Calculations using the model described by eq. (1) indicate that a 2 to 3 fold difference in apparent K_d for two Hfq variants would be difficult to detect, although an eight-fold difference should be observable.

Another factor worth noting is that the gel shift experiment is a pseudo equilibrium method. During the time RNA-protein complexes enter the gel and separate from free RNA and protein the low ionic strength running buffer is exchanging with the 0.5 M Na^+ loading buffer lowering the ionic strength. The electrophoresis process in the gel also results in dissociation of the relatively weak Hfq-$DsrA_{DII}$ complexes as was apparent by the distributed intensity between the free RNA band and the complex bands, and the lack of well defined complex bands for mutant Hfqs using the 2 nM RNA concentration employed to evaluate binding affinities. Formation of $DsrA_{DII}$-Hfq complexes under the gel conditions appears to be more sensitive to the Hfq mutations than the solution conditions used in the anisotropy measurements. The results suggest a need to explore Hfq-$DsrA_{DII}$ interactions in lower ionic strength solutions and to compare dissociation and association rates of wild type and mutant Hfq for $DsrA_{DII}$.

Appendix I

```
B.anthra-1   QFLNQLRKENT-FVTLYLLNGFQLRGLIKGFDNFTVLLETEGKQQLIYKHAISTFVP
B.cereus-1   QFLNQLRKENT-FVTLYLLNGFQLRGLIKGFDNFTVLLETEGKQQLIYKHAISTFVP
B.subtilis   QFLNQIRKENT-YVTVFLLNGFQLRGQVKGFDNFTVLLESEGKQQLIYKHAISTFAP
B.halodura   HFLNQLRKENI-PVTVFLLNGFQLRGLVKGFDNFTVILETEGKQQLVYKHAISTFAP
G.stearoth   QFLNQLRKEGI-QVTVFLLNGFQLRGYIKGFDNFTVLLEVQGKQQLIYKHAISTFAP
O.iheyensi   QYLNQLRKNHI-SVTVFLTNGFQLRGLVKAFDNFTVLLETDGKQQLIFKHAISTFSP
L.innocua    YYLNQLRKEKI-LATVFLTNGFQLRGRVVSFDNFTVLLDVEGKQQLVFKHAISTFSP
L.monocyto   YYLNQLRKEKI-LATVFLTNGFQLRGRVVSFDNFTVLLDVEGKQQLVFKHAISTFSP
C.perfring   IFLNNARKERI-PVTIFLVNGVQLKGIVKGFDSFTVVLDSDGKQQLVYKHAISTVSP
C.acetobut   IFLNSARKNKI-PVAIHLTNGFQMRGSVKGFDSFTVILESDGKQMMIYKHAVSTITP
T.tengcong   IFLNQVRKEHV-PVTVYLINGFQLKGTVKGFDNFTVVLESESKQLLIYKHAISTISP
B.anthra-2   ELYKQIKEEKG-IVTIFLKSGVRIVGEIVAIDKFTVLMLVDGKQQLIYKQAISTIMK
B.cereus-2   ELYKQIKEEKG-TVTIFLKSGVRIIGEVVGVDKFTLLILVDGKQQLIYKQAISTIMK
B.anthra-3   DFYNKLIEEQR-LVTIFLINGVRVPGIIIAVDKFSVLVSSHGKQQFIYKHAISTVSL
S.aureus-1   KALENFKANQT-EVTVFFLNGFQMKGVIEEYDKYVVSLNSQGKQHLIYKHAISTYTV
S.epider-1   QALENFKSEKT-EVTIFFLNGFQMKGVVENYDKYVVSLNSQGKQHLIYKHAISTFTV
B.thurin-2   QLLQEAFQKRK-DITLILLKGLHVKGIIRGFDTYSVLIEVEGKQQLVYKHAISTIRF
B.thurin-1   HLLQEAFQKKK-DITLILLKGLHIKGIITGYDTFSILIEYEGKQQLVYKHAISTIRF
L.interrog   QLLNTARKDKL-DLTIYLLNGVPLKGKVVSFDNFTIVLEQENKQSLVYKHAISTIIP
T.maritima   RFLNHLRVNKI-EVKVYLVNGFQTKGFIRSFDSYTVLLESGNQQSLIYKHAISTIIP
E.coli       PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
S.sonnei     PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
S.flexneri   PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
S.dysenter   PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
S.paratyph   PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
S.enterica   PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
S.typhimur   PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
S.bongori    PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
K.pneumoni   PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
Y.pestis     PFLNALRRERV-PVSIYLVNGIKLQGQVESFDQFVILLKNT-VSQMVYKHAISTVVP
Y.enteroco   PFLNALRRERV-PVSIYLVNGIKLQGQVESFDQFVILLKNT-VSQMVYKHAISTVVP
E.chrysant   PFLNALRRKRV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
P.carotovo   PFLNALRRERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
A.actinomy   PYLNALRRERI-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VNQMVYKHAISTVVP
H.somnus     PYLNALRRERI-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VNQMVYKHAISTVVP
A.pleuropn   PYLNALRRERI-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
H.influenz   PYLNALRRERI-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VNQMVYKHAISTVVP
H.ducreyi    PYLNALRRERI-PVSIYLVNGIKLQGQIESFDQFIILLKNT-VSQMVYKHAISTVVP
P.multocid   PYLNALRRERI-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VNQMVYKHAISTVVP
P.profundu   PFLNALRRERI-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VNQMVYKHAISTVIR
V.parahaem   PFLNALRRERI-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VNQMVYKHAISTVVP
V.cholerae   PFLNALRRERI-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VNQMVYKHAISTVVP
V.vulnific   PFLNALRRERI-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VNQMVYKHAISTVVP
S.oneidens   PFLNALRRERV-PVSIYLVNGIKLQGQVESFDQFVILLKNT-VSQMVYKHAISTVVP
S.putrefas   PFLNALRRERV-PVSIYLVNGIKLQGQVESFDQFVILLKNT-VSQMVYKHAISTVVP
W.brevipal   PFLNTLRRERI-PVSIYLVNGIKLQGYIESFDQFVILLKNS-ISQMIYKHAISTVVP
C.psychroe   PFLNALRRDRI-PVAIYLVNGIKLQGQVESFDQFVILLKNT-VSQMVYKHAISTVVP
X.axonopod   PFLNALRRERV-PVSVYLVNGIKLQGTIESFDQFVVLLRNT-VSQMVYKHAISTVVP
X.campestr   PFLNALRRERV-PVSVYLVNGIKLQGTIESFDQFVVLLRNT-VSQMVYKHAISTVVP
X.fastidio   PFLNALRRERV-PVSIYLVNGIKLQGTIESFDQFVVLLRNT-VSQMVYKHAISTVVP
P.fluoresc   PYLNTLRKEKV-GVSIYLVNGIKLQGTIESFDQFVILLKNT-VSQMVYKHAISTVVP
P.syringae   PYLNTLRKEKV-GVSIYLVNGIKLQGTIESFDQFVILLKNT-VSQMVYKHAISTVVP
```

```
P.putida     PYLNTLRKEKV-PVSIYLVNGIKLQGSIESFDQFVVLLKNT-VSQMVYKHAISTVVP
P.aerugino   PYLNTLRKERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
A.vineland   PYLNTLRKERV-PVSIYLVNGIKLQGQIESFDQFVILLKNT-VSQMVYKHAISTVVP
M.degradan   PYLNVLRKERI-PVSIYLVNGIKLQGQVESFDQFVVLLKNT-VSQMVYKHAISTVVP
B.parapert   PFLNTLRKEHV-PVSIYLVNGIKLQGQIESFDQYVVLLRNT-VTQMVYKHAISTVVP
B.pertussi   PFLNTLRKEHV-PVSIYLVNGIKLQGQIESFDQYVVLLRNT-VTQMVYKHAISTVVP
B.bronchis   PFLNTLRKEHV-PVSIYLVNGIKLQGQIESFDQYVVLLRNT-VTQMVYKHAISTVVP
N.meningit   PFLNALRKEHV-PVSIYLVNGIKLQGQVESFDQYVVLLRNTSVTQMVYKHAISTIVP
N.gonorrho   PFLNALRKEHV-PVSIYLVNGIKLQGQVESFDQYVVLLRNTSVTQMVYKHAISTIVP
B.pseudo-1   PFLNALRKEHV-PVSIYLVNGIKLQGNIESFDQYVVLLRNT-VTQMVYKHAISTVVP
B.mallei-1   PFLNALRKEHV-PVSIYLVNGIKLQGNIESFDQYVVLLRNT-VTQMVYKHAISTVVP
B.cepaci-1   PFLNALRKEHV-PVSIYLVNGIKLQGNIESFDQYVVLLRNT-VTQMVYKHAISTVVP
B.fungor-1   PFLNALRKEHV-PVSIYLVNGIKLQGNIESFDQYVVLLRNT-VTQMVYKHAISTVVP
R.solanace   PFLNALRKEHV-PVSIYLVNGIKLQGNIESFDQYVVLLRNT-VTQMVYKHAISTVVP
R.metall-1   PFLNALRKEHV-PVSIYLVNGIKLQGNIESFDQYVVLLRNT-VTQMVYKHAISTVVP
N.europaea   PFLNILRKERI-PVSIYLVNGIKLQGQIDSFDQYVVLLKNS-VTQMVYKHAISTIVP
A.ferrooxi   PFLNLLRKEHV-PVAIYLVNGIKLQGFVESFDQFVVLLRNN-VSQMIYKHAISTVVP
M.capsulat   PFLNTLRKEHV-PVSIYLVNGIKLQGKVDSFDQYVIMLKNT-VSQMVYKHAISTIVP
D.nodosus    PYLNALRKERV-PVSIYLVNGIKLQGQIESFDAFVILLRNN-ISQMVYKHAVSTIVP
Magnetococ   PFLNTLRREKV-PVTVFLVNGIKLQGMITSFDNYCLLLKNS-VTQLVFKHAISTVMP
L.pneumoph   PFLNELRKEKV-PVSVFLVNGIKLHGIIDSFDQYVVMLKNS-ITQMVYKHAISTVVP
B.pseudo-2   DFINAARKERK-RVEIYLVNGIRLTGCIESFDQYLVMLRTPVGLQGIYKRAISTIQL
B.mallei-2   DFINAARKERK-RVEIYLVNGIRLTGCIESFDQYLVMLRTPVGLQGIYKRAISTIQL
B.cepaci-2   DFINSARKERK-RVEIYLVNGIRLTGCIESFDQYLVMLRTPVGLQGIYKRAISTIQL
B.fungor-2   DFMNAARKERK-RVEIYLVNGIRLTGCIESFDQYLVMLRTPVGLQGIYKRAISTIQL
A.aeolicus   SFLNTARKKRV-KVSVYLVNGVRLQGRIRSFDLFTILLEDGKQQTLVYKHAITTIVP
B.melitens   LFLNSVRKQKI-SLTIFLINGVKLTGIVTSFDNFCVLLRRDGHSQLVYKHAISTIMP
B.suis       LFLNSVRKQKI-SLTIFLINGVKLTGIVTSFDNFCVLLRRDGHSQLVYKHAISTIMP
M.loti       LFLNSVRKSKN-PLTIFLINGVKLTGVVTSFDNFCVLLRRDGHSQLVYKHAISTIMP
A.tumefaci   LFLNTVRKQKI-SLTIFLINGVKLTGVVTSFDNFCVLLRRDGHSQLVYKHAISTIMP
R.legumino   LFLNTVRKQKI-SLTIFLINGVKLTGVVTSFDNFCVLLRRDGHSQLVYKHAISTIMP
S.meliloti   LFLNTVRKQKI-SLTIFLINGVKLTGVVTSFDNFCVLLRRDGHSQLVYKHAISTIMP
B.japonicu   TFLNHVRKTKT-PLTIFLVNGVKLQGIVTWFDNFCLLLRRDGHSQLVYKHAISTIMP
R.palustri   TFLNHVRKTKT-PLTIFLVNGVKLQGIVTWFDNFCLLLRRDGHSQLVYKHAISTIMP
M.magnet-2   TFLNHVRKNKI-PLTIFLVNGVKLQGVVTWFDNFCVLLRRDGHSQLVYKHAISTIMP
C.crescent   TFLNSVRKSKT-PLTIFLVNGVKLQGVVSWFDNFCVLLRRDGQSQLVYKHAISTIMP
R.sphaeroi   AFLNHVRKAKV-PVTIFLINGVKLQGVITWFDNFCVLLRRDGQSQLVYKHAISTIMP
S.pomeroyi   AFLNHVRKTKV-PVTIFLINGVKLQGVITWFDNFCVLLRRDGQSQLVYKHAISTIMP
M.magnet-1   VFLNYIRKNKT-PVTIFLVNGVKLQGIVTWFDNFSVLLRRDGHTQLVYKHAISTVMP
R.rubrum     VFLNYIRKNKA-PVTIFLVNGVKLQGIVTWFDNFSLLLRRDGHTQLVYKHAISTIMP
H.neptuni    TFLNAVRKSRT-PLTVFLVNGVKLQGVVTWFDNFCILLRGDGRPQLVYKHAISTIAP
N.aromat-1   LFLNHLRKNKI-PVTMFLVKGVKLQGIVTWFDNFSILLRRDGQSQLVYKHAISTIMP
N.aromat-2   VFLSSVRDSGV-QVTMFLVNGVMLQGKVASYDLFCMLLEREGYVQLAYKHAVSTIQP
C.hydroge*   AFLNQVRKENVGGLPIFLFNGFQLKRFCKRFLTFLLVFWESEGKHMIYKHAISTIIP
R.albus      VFLNQARKEQV-MVKFILMNGYQFKGIVKAFDSY-VVFLDCEGKNVVYKHAISTIVP
G.sulfurre   QYLNQSRKERI-KVAVRLMSGEKLEGYIKSFDNFSVLMEIQGDM-LIYKHAITSITS
R.metall-2   LHYAAHGASRN-AVIVHLSNGTRLTGVVLASDNYMVLLGQSAEDTLIYKRAITVVTP
M.jannasc*   VIFEYARRLNGKKVKIFLRNGEVLDAEVTGVSNYEIMVKVGDRNLLVFKHAIDYIEY
S.aureus-2   YILVKLTLTNN-KILIGKVID--FDDKVDNFDGYN-SIEIDTGRTYDISENKIKTIV
S.epider-2   YLIVKLTLTNN-KVLVGKVTG--FDDKYDNFDGFN-SIEIDTGSLYDITENKIKSIV
F.acidarm*   GFIFGKNLIGK-TMNITLLNGEILSGVLRGFGQYDILLE-SGGKVILMKSGIVKIEV
T.volcani*   SVEINTGLEGK-VLTISLLNGRIEAGKLKVAGQYFLEIEGANGRLIIAKSAIVTVSV
P.horikos*   VYQDIFGVPLGSEVNIALQNKLQISGVLKGYKEGFLLVQRGNSLVLVNPNEIAYISV
```

Figure 3.24 Multiple alignments of All Hfq sequences (* indicates it is in archaea)

Appendix II

The exact evolutionary relationship between archaea and eukarytic Sm is still not clear, but the most generally accepted view is that the two proteins evolve from a common ancestor that diverged from bacterial Hfq (Zhang *et al.,* 2002, Moll *et al*., 2002). It has been found that Hfq and Sm protein have many similar functions, such as serving as a host factor for RNA virus replication, participating in RNA degradation and facilitating RNA duplex formation. The function of this protein family appears to have been conserved during evolution even though its oligomeric form changed from hexamer to heptmer.

To examine that, a multiple alignment of the archaea Sm protein was built up according the genome blast search (shown as Fig3.25A). Each archaea genome contains either one or two copy of Sm protein, termed as Sm-1 and Sm-2. In all the archaea Sm-1 proteins, the three residues which are correspond to the Leu12, Phe39 and Phe42 in Hfq copy A are highly conserved, but with His the dominating residue at in position 37 (corresponding to Phe at position 42 in Hfq). In contrast, the highly conserved positively charged Arg residue at position 16 and 17 were not observed in Sm proteins. A previous study on *Methanobacteriaum thermoautotrophicum* Smα protein also noticed that the adjacent Leu9 and Phe34 (corresponding to Leu12 and Phe39 in Hfq) are a highly conserved hydrophobic pair in archea, and suggested that this region may be involved in protein-protein interaction. Recently however Thore et al (Thore *et al.,* 2003) showed Tyr34 in a crystal structure of *Pyrococcus abyssi* Smα protein can form contacts with RNA. Mutating this residue to Val reduced the binding of this Smα protein to the Sm

consensus RNA oligomer. Figure 3.25B also shows that the covariance between Leu9 and Phe/Tyr34 is observed in some Sm proteins such as *M. thermoautotrophicum* Smβ, *D.melanogaster* SmF*, A. thaliana* Lsm6, and *C.elegans* Lsm2. If one examines the distal surface of the Sm protein, the counterpart of the three polyA binding residuesTyr25, Ile30 and Lys31 in Hfq are not found in the archeal Lsm proteins (figure 3.25A). We note that almost all RNAs known to interact with Sm proteins are U-rich RNA. No A-rich RNA has been reported.

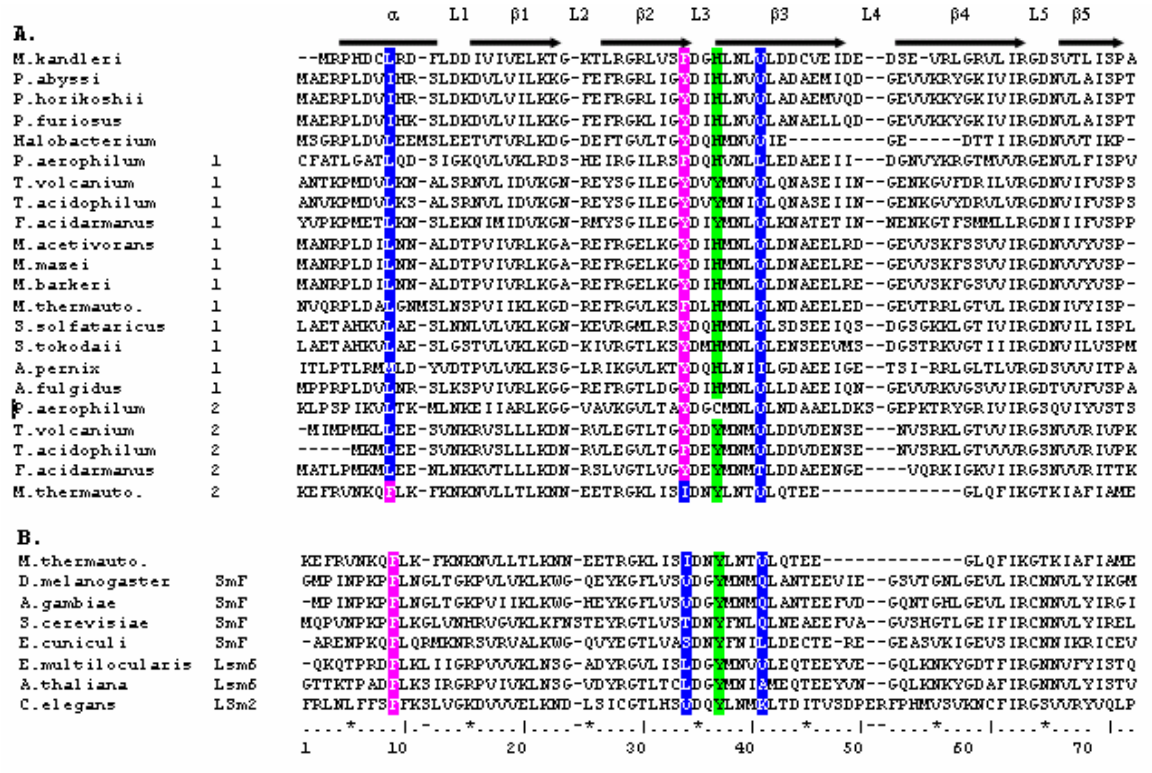


Figure 3.25. The multiple alignment of Sm/Lsm protein

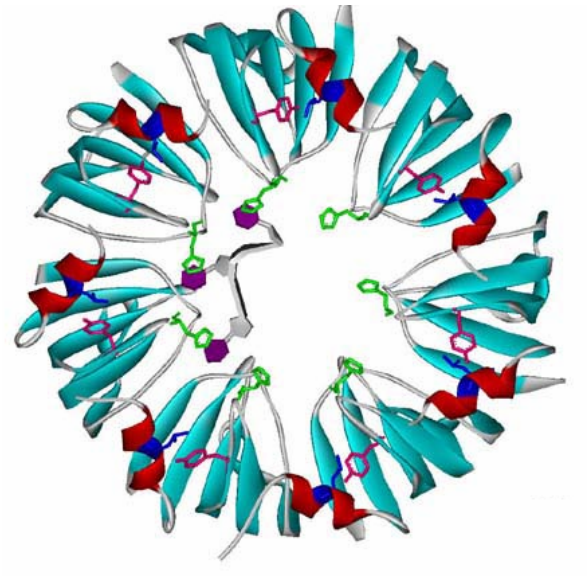


Figure 3.26 The 3-D structure of *A. fulgidus* Sm1.The aromatic residues in position 34 and 37 of archeal Sm were shown in green and pink respectively, and the aliphatic residue which covaries with the aromatic residue 34 was shown in blue.

References

Altuvia, S., Zhang, A., Argaman, L., Tiwari, A. & Storz G., (1998) The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. *EMBO J*. **17**, 6069-6075.

Antson,A.A., Dodson,E.J., Dodson,G., Greaves,R.B., Chen, X-P. and Gollnick,P. (1999) Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature*, **401**, 235–424.

Brescia CC, Mikulecky PJ, Feig AL, Sledjeski DD. (2003) Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. *RNA*. **9**:33-43.

Carmichael C.G. and Weber K., (1975) The host factor required for RNA phage Qβ RNA replication *in vitro Journal Biological Chemistry* **250,** 3607-3612.

Collins, B.M., Harrop S.J., Kornfeld, G.D., Dawes, I.W., Curmi, P.M & Mabbutt, B.C., (2001) Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs. *J.Mol. Biol*, **309**, 915-923.

de Haseth, P.L. and Uhlenbeck, O.C. (1980a) Interaction of *Escherichia coli* host factor protein with Q beta ribonucleic acid. *Biochemistry* **19**: 6146–6151.

de Haseth, P.L. and Uhlenbeck, O.C. (1980b) Interaction of *Escherichia coli* host factor protein with oligoriboadenylates. *Biochemistry* **19**: 6138–6146.

Draper D.E., (1999*)* Themes in RNA-protein recognition. *J. Mol. Biol*. **293**, 255-270.

Folichon M, Allemand F, Regnier P, Hajnsdorf E. (2005) Stimulation of poly(A) synthesis by Escherichia coli poly(A)polymerase I is correlated with Hfq binding to poly(A) tails. *FEBS J*. **272**:454-63.

Gill, SC, von Hippel, PH (1989) Calculation of. protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* **182**, 319-326.

Hajnsdorf E, Regnier P. (2000) Host factor Hfq of Escherichia coli stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proc Natl Acad Sci U S A*. **97**:1501-5.

Jones S., Daley, D.T., Luscomebe, N.M. & Berman H.M., (2001), Protein-RNA interactions: a structural analysis. *Nucleic Acids Res*. **29**, 943-954.

Lease RA, Woodson SA. (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J Mol Biol*. 344:1211-23.

Le Derout J, Folichon M, Briani F, Deho G, Regnier P, Hajnsdorf E.(2003) Hfq affects the length and the frequency of short oligo(A) tails at the 3' end of Escherichia coli rpsO mRNAs. *Nucleic Acids Res*. **31**:4017-23.

LeTilly V, Royer CA.(1993) Fluorescence anisotropy assays implicate protein-protein interactions in regulating trp repressor DNA binding. *Biochemistry*. **32**:7753-8.

Lundblad JR, Laurance M, Goodman RH. (1996) Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Mol Endocrinol*. **10**:607-12.

Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, Feig AL.(2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol.*, **11**:1206-14.

Miranda G, Schuppli D, Barrera I, Hausherr C, Sogo JM, Weber H. (1997) Recognition of bacteriophage Qbeta plus strand RNA as a template by Q beta replicase: role of RNA interactions mediated by ribosomal proteins S1 and host factor. J Mol Biol., **267**:1089-103.

Moll, I., Leitsch, D., Steinhauser, T., and Blasi, U. (2003b) RNA chaperone activity of the Sm-like Hfq protein. *EMBO Rep* **4**: 284–289.

Mol1er, T., Franch, T., Hojjrup, P., Keene, D.R., Bichinger, H.P., Brennan, R.G., and Valentin-Hansen, P. (2002a) Hfq: a bacterial Sm-like protein that mediates RNA–RNA interaction. *Mol Cell* **9**: 23 30.

Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, Moll I, Kachalova G, Yokoyama S, Vassylyev D, Garber M, Nikonov S. (2005) Structure of Pseudomonas aeruginosa Hfq protein. *Acta Crystallogr D Biol Crystallogr*. **61**,141-6

Sauter C., Basquin J., Suck D., (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. *Nucleic Acids Res*., **31**, 4091-4098.

Schumacher, M.A., Pearson, R.F., Moller, T., Valentin-Hansen, P. & Brennan, R.G., (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J*. **21**, 3546-3556.

Sonnleitner E, Napetschnig J, Afonyushkin T, Ecker K, Vecerek B, Moll I, Kaberdin VR, Blasi U. (2004) Functional effects of variants of the RNA chaperone Hfq. *Biochem Biophys Res Commun*. **323**:1017-23

Storz G, Opdyke JA, Zhang A. (2004) Controlling mRNA stability and translation with small, noncoding RNAs. Curr Opin Microbiol. **7**:140-144.

Sun, X., Zhulin, I., Wartell, M.W., (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res*. **30,** 3662-3667.

Thore S, Mayer C, Sauter C, Weeks S, Suck D. (2003) Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. J Biol Chem **278**(2):1239-47.

Valentin-Hansen P, Eriksen M, Udesen C. (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. Mol Microbiol. 51(6):1525-33.

Zhang, A.X., Wassarman, K.M., Ortega, J., Steven, A.C., and Storz, G. (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell* **9**: 11–22.

.

CHAPTER IV

A STRUCTURE TRANSITION INVOLVING DIMERIZATION OF 79 NT RNA

Introduction

RNA represents a unique class of molecules in cell due to its ability to carry information and have catalytic activity (Joyce, 1996). Both functions depend on interactions of RNA molecules with other molecules. RNA-RNA interactions are involved in many biological processes, such as pre-mRNA splicing (Green, 1986), RNA editing (Koslowsky, 2004), gene expression regulated by RNAi or antisense RNA (Lavorgna *et al.,* 2004), as well as the recent identified ribo-regulation by specific small RNAs (Storz *et al.*, 2004).

RNA dimerization is a special case of this kind of interaction. It has been observed to occur for retroviral RNA genomes where two homologous RNA molecules form a complex (Paillart *et al*., 1996). This dimerization complex is essential for viral function and is initiated at a stem-loop structure named the dimerization initiation site (DIS). The DIS of HIV-1 is a highly conserved structure with a self-complementary loop sequence that forms a loop-loop 'kissing' complex.

We have found a different mechanism of RNA dimer formation occurring for an *in vitro* transcribed 79 nt RNA. This dimer is enhanced in presence of Mg^{2+} and low temperature. RNA structure prediction (using the program Mfold 3.1) indicates that the first 21 nt from the 5' end of the 79 mer may be critical for dimerization. A transition of

short intramolecular stem-loops to an intermolecular duplex appears to be responsible. Several methods including an RNase H assay, circular dichroism spectroscopy and sedimentation equilibrium centrifugation are consistent with this prediction.

Materials and Methods

Plasmid construction

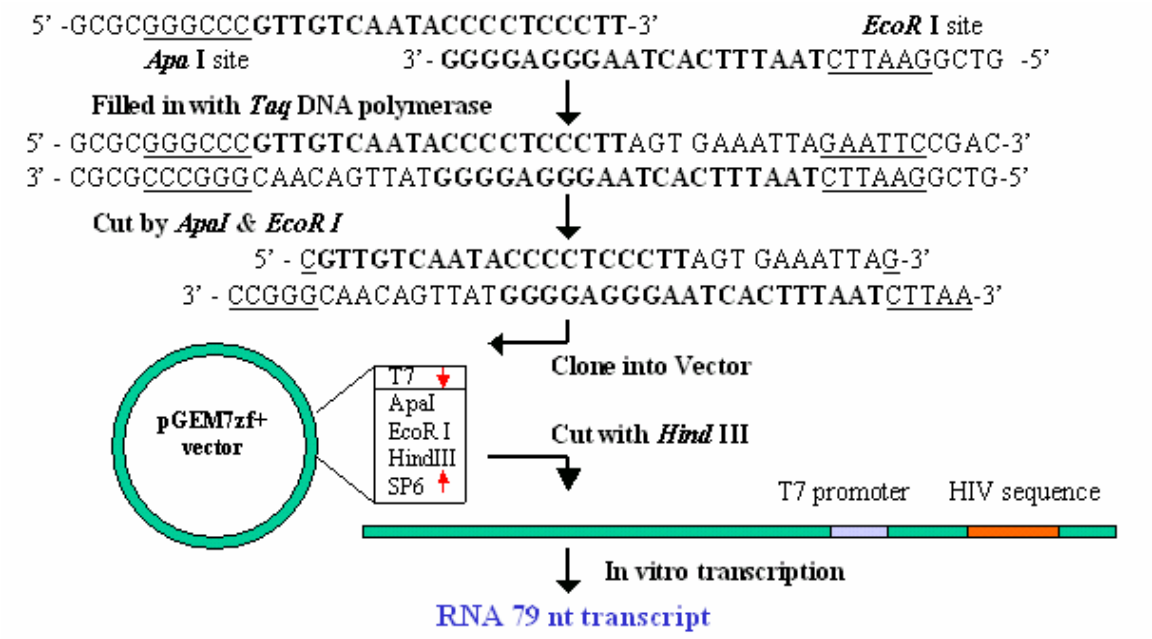


Figure 4.1 Construction of the modified pGEM7Zf(+)

The pGEM7Zf(+) plasmid from Promega Inc. was modified to provide a template for RNA transcription. A 30 nt sequence from the HIV-1 reverse transcriptase gene containing 10 consecutive pyrimidines was inserted into the plasmid multiple cloning site. We refer to this new plasmid as pN1. Two partially complementary DNA oligomers (5'-GCGCGGGCCCGTTGTCAATACCCCTCCCTT-3' and 5'-GTCGGAATTCTAAT

TTCACTAAGGGAGGGG-3') were hybridized and filled in by Taq DNA Polymerase. The underlined bases are from HIV-1. After digestion with *ApaI* and *EcoRI* the duplex oligomer was ligated into pGEM7Zf(+) between *ApaI* and *EcoRI* restriction sites downstream of a T7 promoter. The plasmid was treated with calf intestinal alkaline phosphatase prior to ligation. The ligated plasmid was transformed into *E coli* JM 109 cells, and colonies selected for ampicillin resistance. The desired construct was identified by *DdeI* digestion. pN1 generates 5 fragments, while the parent plasmid produce 4 *Dde I* bands.

***In vitro* transcription**

79 nt RNA transcripts were produced using *HindIII*-cut pN1 plasmid as the template for T7 RNA Polymerase. A 1 ml transcription reaction contained 10 mM dithiothreitol, 500 μM of ATP, CTP, GTP and UTP, 1X transcription buffer (50 mM Tris-HCl, pH7.5, 10 mM NaCl, 6mM $MgCl_2$ and 2mM spermidine), 25 units of Prime RNA inhibitor (5 Prime-3 Prime, Inc), 0.1 mg of BSA (1mg/ml), 120 units of T7 RNA polymerase, and 100 μg of *Hind III* digested pN1 plasmid DNA. Transcription reactions were carried out in a 37 °C water bath for 3 hours. Afterwards, the pN1 plasmid template was separated from the remaining reaction components with a Centricon-100 concentrator (Amicon). The filtrate was collected and transferred to a Centricon-10 concentrator and centrifuged at 1000x g to concentrate the RNA and to remove smaller molecules. The last step was repeated several times. 50 nt RNA transcripts were produced using *EcoRI*-cut pN1 plasmid as the template. RNA concentration was

estimated from its molecular weight and UV absorbance at 260 nm. It was assumed that 1.0 OD at 25 °C equals 40 μg/ml. RNA was also characterized by PAGE.

Temperature Gradient Gel Electrophoresis (TGGE) of RNA

The method of TGGE was applied for the structural analysis of the RNA transcripts. The apparatus for running the vertical temperature-gradient gel was described previously (Wartell, *et al.*, 1990). Two different electrophoresis buffers were used. The 1XTBE buffer contained 90 mM Tris-borate and 5 mM Na_2EDTA. The 1X TBM buffer contained 90 mM Tris-borate and 5 mM $MgCl_2$. To set up the temperature gradient in the gel, two aluminum heating blocks were sandwiched the gel glass plates. Two water circulators were connect with the aluminum heating blocks and established the temperature gradient perpendicular to the electric field. The temperatures were measured in the gels with a needle-like thermo-couple probe (TMTSS-020-6, Omega Inc.) connected to a digital thermometer (MDSD-465, Omega Inc, accuracy estimated as ±0.1°C). Measurements were taken at two positions at the end of each electrophoresis with the electricity being turned off while keeping the water circulator running. The temperature gradient was linear and uniform within the region covered by the heating blocks. The gels were stained with ethidium bromide and photographed. The temperature probes were usually inserted between the edge of the gel and the gel spacer and thus was very easy to be observed in the photograph and provided a temperature scale.

Sedimentation equilibrium experiments

A sedimentation equilibrium method was employed to measure the molecular weight of the RNA in different temperature and buffer condition. The method with minor modifications was similar to the procedure previously described by Dripps and Wartell, 1987, A Ti50 rotor was employed in a Beckman preparative ultracentrifuge. Teflon rods were machined into the shape of a rotor centrifuge tube. A hole was drilled into the top of each unit in order to hold a 250 µl round bottom tube (Fisher Scientific). Three ^{32}P – labeled marker DNAs and one ^{32}P –labeled sample RNA were run at 20K rpm for about 20 hours. The solvents are either TBE or TBM buffer containing 5 µg/µl ficol to stabilized convection. The solvents also contain Prime RNase Inhibitor (5 Prime-3 Prime, Inc) to prevent RNA degradation. Each tube contains 100 µl of sample. After centrifugation, the uppermost 40 µl and the remaining 60 µl at the bottom were removed from each tube. The amount of the sample in the solution was then determined by using scintillation counting. The fraction of sample remaining in the upper 40 µl after centrifugation, F, was determined for each sample. The unknown molecular weight of the RNA sample was evaluated from a plot of log F vs. the molecular weight of the reference DNAs.

CD spectroscopy

Circular dichroism spectra were obtained on a Jasco J-720 spectropolarimeter equipped with a MESLAB RTE-111 circulating bath. Each spectrum was an average of four scans and was collected at a rate of 50 nm/min with a bandwidth of 0.5 nm and sampling wavelength of 1 nm. The CD spectra were recorded from 320 to 200 nm at four different temperatures, namely 20, 40, 60, 80 °C and normalized by subtraction of the

background scan with buffer. The data was processed on a PC computer using Windows-based software supplied by the manufacturer (JASCO Inc.)

RNase H Assay

Two 40 µl 'master mixture' solutions correspondent to the 79mer monomer and dimer were prepared by mixing the RNA and *E coli* RNase H1 (USB), and preincubated for 15 min at 25 °C in RNase H buffer with 0.15 unit of prime RNase inhibitor. 7 µl from this mixture was added to each reaction tube with 3 µl of a given DNA. Each reaction was incubated at 25 °C for 30 min. Sample were heated at 90 °C for 2-3 min and then quickly cool on ice. They were mixed with ficol loading buffer, and run immediately into a 12% native PAGE. 0.5 unit of RNase H was used per reaction, and the RNA and DNA concentration were 1.5 and 9 µM, respectively. Nucleic acid bands were visualized by staining the gel with ethidium bromide followed by UV-induced fluorescence. Gel image was captured with a video gel documentation system. Exposure times were adjusted to produce images of the uncut RNA band in a linear intensity range. The relative intensities of the RNA bands were determined using the NIH Gel image analysis program. Percentages were reproducible within ± 5%.

Hybridization Assay

The 50 mer RNA corresponding to the first 50 nt of the 79 mer was transcribed by using *EcoR I*-cut pN 1 plasmid as the template instead of *Hind III*-cut pN1 for the 79 mer. The 47 mer corresponding to the last 47 nt of the 79 mer was produced by a site-specific RNase H cleavage of the 79 mer under the help of a 2'-O'-methyl -RNA/DNA chimera

(5'-$C^mU^mA^mA^mG^mGGATG^mG^mG^mU^m$-3') (Inour *et al*., 1988). After gel separation and purification, the resulting 47 mer as well as the 50 mer and 79 mer was mixed together respectively in an equal-molar ratio, then heated to 85 °C for 3 minutes and quick cooled on ice for one hour. These three mixed samples were loaded in a 12 % PAGE gel with the individual RNAs with or without the heating-cooling process.

Results and Discussion

Non denaturing PAGE studies of 79 nt RNA

The purified RNA transcript showed different band patterns in different conditions. In a 12% denaturing PAGE containing 8M urea or in a 12% non-denaturing PAGE with RNA being preheated to 85 °C for 3 minutes prior to loading, only one main band was detected. However, without pre-heating the RNA sample, two bands were seen in a 12% native PAGE at room temperature with 1x TBE as gel running buffer with the faster moving band the dominant band (Figure 4.2B). When the sample was run in 1xTBM buffer containing 89mM Tris-borate and 5 mM $MgCl_2$, the two RNA bands had similar intensities (Figure 4.3B). The gel pattern of the purified RNA transcription product indicates that one RNA molecule is produced after the transcription and purification procedure, and different bands on the gel are from the same RNA molecules

Figure 4.2 A shows the RNA gel pattern run at 4 °C in TBE gel buffer (89 mM Tris-borate and 2mM EDTA, pH 8.0). The RNA was preheated to 90 °C for 3 minutes and quickly put on ice for a few minutes and then maintained at 4°C for up to 2 hours. The gel pattern shows that the slower moving group I bands disappear. The faster moving

group II bands display four or five closely spaced bands. This suggests that RNA can refold into different secondary/tertiary structures in a short period of time (within 10 mins) at room temperature in TBE, only one RNA bands is observed for up to 2 hours after the heating process.

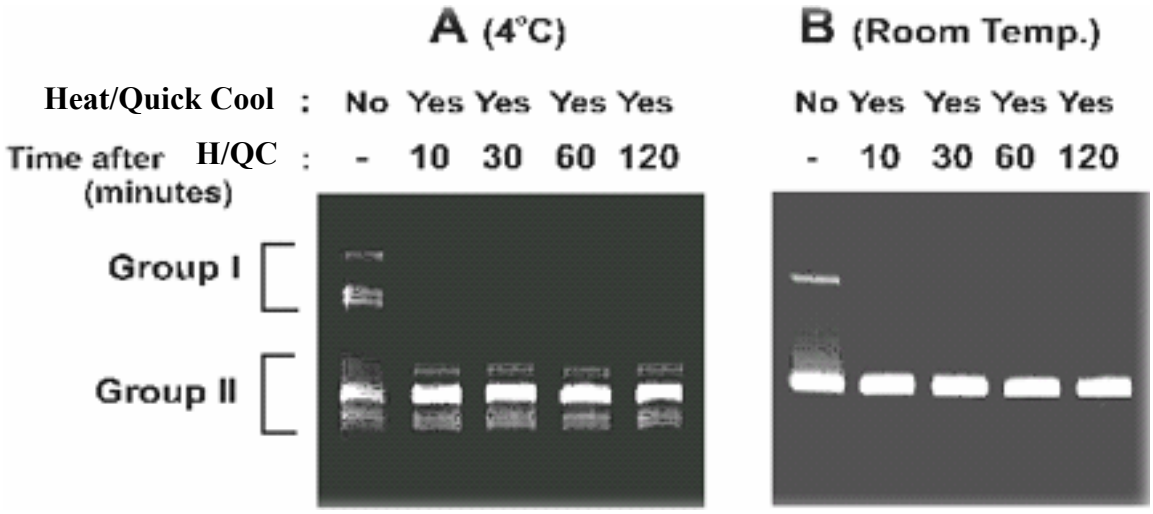


Figure 4.2 Refolding of the RNA molecules after heating assayed by 12% PAGE in TBE buffer. (A) at 4 °C, RNA 79 mer has multiple conformations, after the heating process, the different RNA structures in Group II can be refolded within 10 min, but refolding to RNA structures in Group I is extremely slow process. (B) At room temp, two bands are observed. After heating, only one RNA is observed for up to 2 hours. (data from Jing Li Thesis 2003)

Figure 4.3A shows the behavior of the 79 nt RNA at 4 °C in TBM gel buffer (89 mM Tris-borate and 5mM $MgCl_2$) following the heating/cooling process. The slower moving RNA band observed after transcription and purification is no longer observed for up to 2 hours after heating and quick cooling the sample to 4 °C. A similar behavior of the slower moving RNA band is observed after heating and cooling to 25 °C. The

reconstitution of the RNA structures corresponding to the slower moving band is a relatively slow process in both the TBE and TBM buffer.

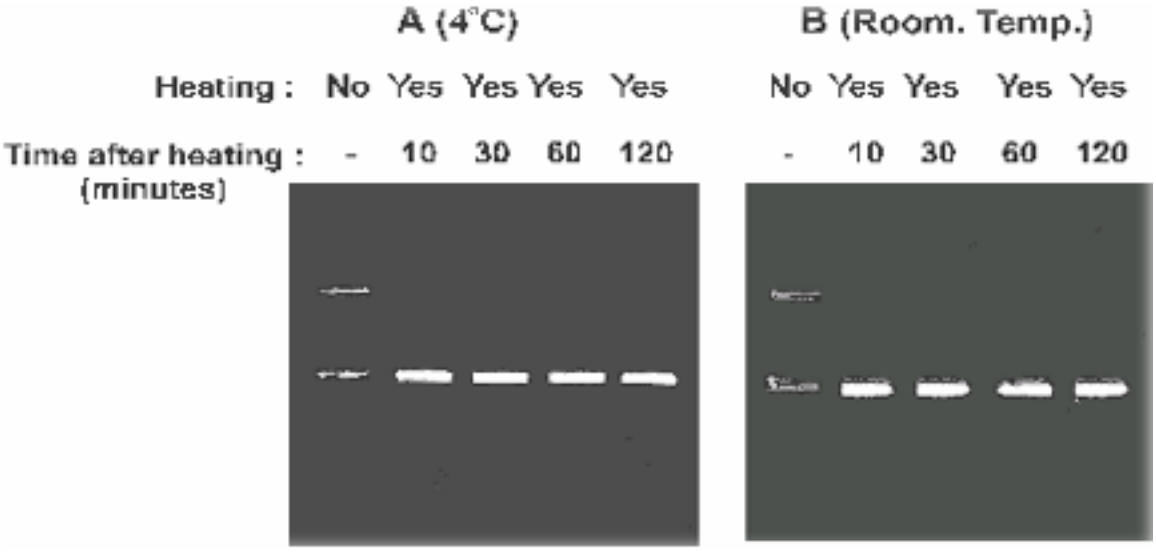


Figure 4.3 Refolding of the RNA molecules assayed by 12% PAGE in TBM gel buffer. (A) At 4 °C, only two single bands are observed. After the heating and cooling process, the slower moving RNA band does not reappear for up to 2 hours. (B) At room temp, again, no reappearance of the slower moving RNA band occurs for at least 2 hours after the heating and cooling step. (data from Jing Li Thesis 2003)

RNA molecules produced by in vitro transcription run off from a *HindIII* site are expected to have small distribution of lengths, the 79 nt RNA is likely to have some length heterogeneity. Another possible interpretation of the group I bands is that polyacrylamide gels run at low temperatures can distinguish small length difference in RNAs in the range of 79 nt which are not detectable at higher temperatures, However the longer/shorter minor transcripts would be expected to broaden the RNA band and it is unclear whether they would form distinctive bands. Also, in the 1x TBM buffer at lower temperature, only one band exists in the group I band at low temperature. Therefore, we

favor the interpretation that the RNA molecules in the group I bands correspond to different secondary structures at lower temperature in the 1xTBE buffer. The predicted free energies of the four most stable 79 nt RNA structures at different temperatures were calculated using Zuker's MFOLD version 3.1 (Zuker *et al.* 1999; Mathews *et al.* 1999). At 5 °C, the free energy values of the four most stable RNA structures are relatively close to each other (within 4%) while at higher temperatures, 25 °C, the most stable low energy structure has a free energy value 6.6% higher than the next most stable structure. Although the predicted free energies are based on parameters evaluated in 1 M NaCl, they suggest a temperature dependence of the RNA's structural states that may qualitatively explain the temperature dependence observed by the PAGE experiments. The multiple group I RNA bands observed with roughly equal intensity at low temperatures might reflect different conformations with similar free energy values. As the temperature increases, the separation in free energy between the most stable conformations and the next lowest free energy conformation increases such that only one conformational state dominates.

TGGE studies of the RNA.

Figure 4.4. A and B show the temperature gradient gel of the RNA in TBE buffer (89 mM Tris-borate and 2mM EDTA, pH 8.0). The result indicates that in the absence of $MgCl_2$, the RNA folds into multiple different structures from 5.8 °C to 20.5 °C. There are 4 major bands at 20 °C. At the lower temperatures the RNA molecule shows the two groups of bands, I and II, observed in Figure 4.2 A. The relative intensities of the individual bands within the fast and slow moving groups are about the same. The slowest

moving bands (the top band in the gel) disappears around 22.1°C, while the other topmost bands disappear at 23.4 °C and 35.4 °C respectively.

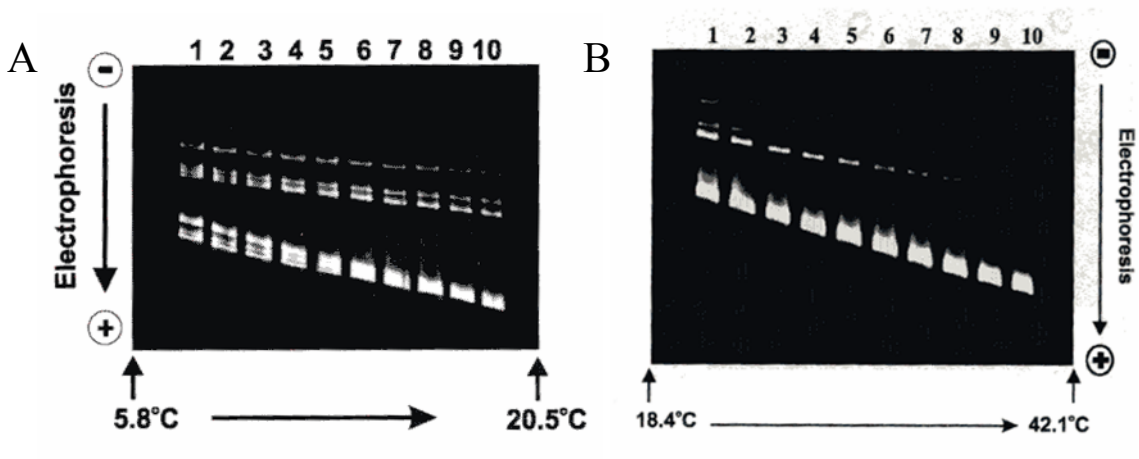


Figure 4.4 Temperature Gradient Gel Electrophoresis of the RNA in TBE buffer.(data from Jing Li Thesis 2003)

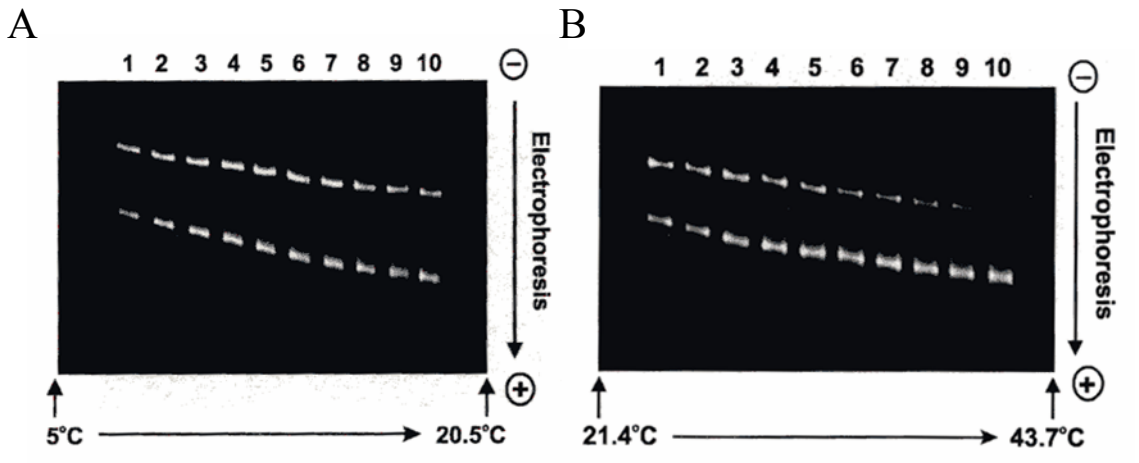


Figure 4.5 Temperature Gradient Gel Electrophoresis of the RNA in TBM buffer.(data from Jing Li Thesis 2003)

Figure 4.5 A and B show a temperature gradient gel of the RNA in TBM buffer (89 mM Tris-borate and 5mM $MgCl_2$). The results indicate that in the presence of 5 mM $MgCl_2$, the RNA folds into only two structures from 5 °C to 39 °C. The two bands have

almost the same intensity until about 30 °C. The slower moving band gradually decreased in intensity with increasing temperature and disappears at 39 °C. A similar outcome is observed using TBC buffer (89 mM Tris-borate and 5mM $CaCl_2$). (data not shown).

The Mg^{2+} or Ca^{2+} - induced conformational change is reversible and is a very fast process once the Mg^{2+} or Ca^{2+} are diluted or removed from the RNA buffer solution. When the RNA that had been dissolved in TBM or TBC buffer but then ran on the gel with TBE buffer, the band pattern was similar to the RNA in TBE buffer only.

Sedimentation equilibrium experiments

The sedimentation equilibrium method was employed to measure the molecular weight of the RNA in different temperature and buffer condition. Three DNA molecules were used as molecular weight references in the sedimentation experiments. One DNA oligomer was 25 bases long with a molecular weight of 7,707, a second DNA oligomer was 47 bases long with a molecular weight of 14,584, and the third molecule was 47 bp DNA duplex with a molecular weight of 28,975. The three DNAs along with the RNA sample were run in separate tubes at 20,000 rpm for 20-22 hours at 5 °C or 28 °C. For both the DNA and RNA samples, the buffers used were either 1xTBM or 1x TBE buffer containing 5 μg/μl ficol to stabilize convection and 0.4 U/μl of Prime RNase inhibitor (5 prime→3 prime, Inc) to prevent RNA degradation. The average molecular weight of the RNA molecule at 28 °C was determined to 1.1 and 1.6 times the monomer molecular weight of the RNA in TBE and TBM buffer respectively, Given the estimated error of these measurements to be ±10%, this indicates the RNA is monomer under the conditions as observed in the gel in 1xTBE buffer and is a mixture of monomer and dimer (or higher

multimer) as observed in the gel in 1xTBM buffer. At 5 °C, the average molecular of the RNA was 1.72 and 1.6 times the RNA monomer for TBE and TBM buffer respectively. The latter values are consistent with the presence of dimers (or higher multimers) in the population of RNAs at the lower temperature in TBE, TBM buffers. This is consistent with the assignment of the group II bands in the gels to dimers or higher multimers. The higher average molecular weight of the RNA in the absence of the Mg^{2+} at 5 °C is consistent with the larger fraction of group II bands observed in the TBE vs. TBM gel. In the 1xTBM buffer, the RNA molecular weight does not change much at 5 °C vs 28 °C in the sedimentation experiment, which is also consistent with the TGGE gel results. One may conclude that the slower moving group II bands observed in the gels are dimer and/or higher order multimers of the RNA.

Predicated structure of 79 mer monomer and Dimer

Both gel electrophoresis and sedimentation equilibrium experiments indicate that the 79 nt RNA can form a monomer and dimer simultaneously and the formation of dimer is enhanced in the presence of Mg^{2+} and low temperature. It remains unclear how the dimer is formed. Does it employ a mechanism similar to the loop-loop kissing complex which occurs in the dimerization of retrovial RNA genome, or a totally different mechanism? To address this question, it is necessary to know the structure of 79mer monomer. The secondary structure with the lowest free energy was predicted using Zuker's MFOLD version 3.1 (Zuker *et al*. 1999; Mathews *et al.* 1999). It implies that the monomer is mainly composed of two stem-loop hairpins flanking a single strand region. This structural feature suggest the possibility of forming a dimer via the kissing complex,

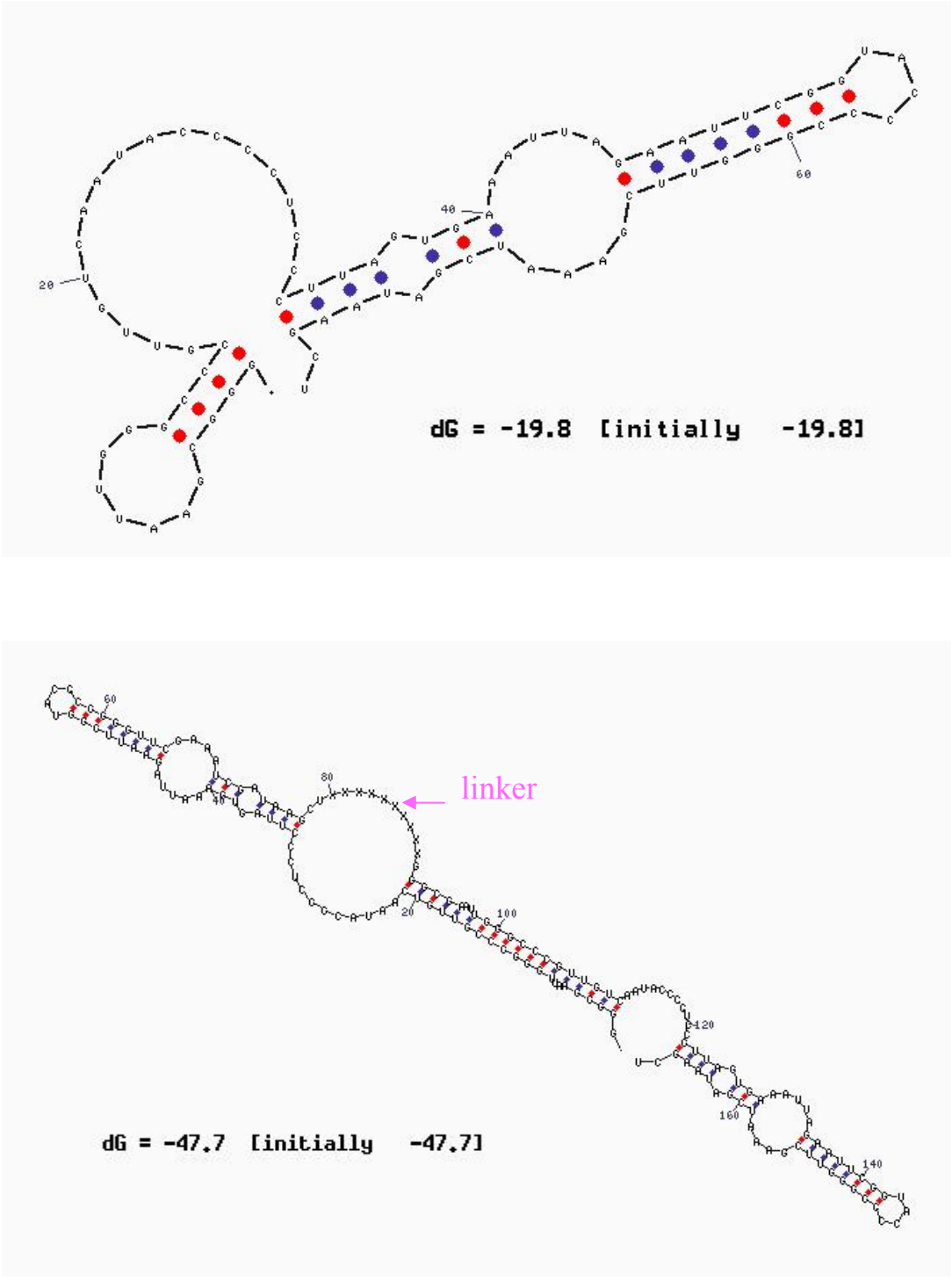


Figure 4.6 The predicted lowest free energy secondary structure of 79 nt RNA monomer and dimer

but sequence examination indicated that only a limited base-pairing could be formed by this fashion which may not energy favorable for dimerization. However, a prediction of the dimer secondary structure may provide some hints on how this intermolecular interaction occurs. To make the prediction, a linker consisted of ten non-identified nucleotides X was used to connect the two monomers in a 5'-3'-5'-3' fashion and then submitted to the MFOLD server. The lowest free energy given by this structure is –47.7 Kcal/mole, about 7.5 Kcal/mole lower than sum of the two individual monomers. So it is energy favorable to form the dimer. It was shown that first 21nt from the 5' end of the 79mer appears to be critical for the dimerization, changing from intramolecular stem–loop to intermolecular duplex, while the second stem-loop remains the same in both monomer and dimer.

Circular dichroism spectra analysis

Circular dichroism spectroscopy is a powerful technique for studying nucleic acid structure. It is more sensitive to the changes in nucleic acid backbone conformation compared with absorbance spectroscopy, and is widely used to monitor the RNA secondary and tertiary folding transition. Previous studies have shown that a model of the secondary structure of an RNA sequence can be evaluated by comparing its CD spectrum with a library of CD spectra of bases and dinucleotides in duplex and single stranded structures (Johnson *et al*., 1991). The fractions of different type of nearest neighbor base pairs in duplex conformations as well as different nucleotides in single stranded conformation in the RNA structure can be estimated. A structural rearrangement in an RNA can be deduced to some extent from the observable change in CD spectrum. For the

79 nt RNA, the predicted structures suggest that the first 21 nt in the 5' end are critical for dimerization. A small stem-loop structure and some single-stranded region was predicted to be replaced by an intramolecular duplex during the transition from monomer to dimmer. This may account for the magnitude difference in their CD spectra at 20 °C. (shown as Figure 4.7A). On the contrary, the shape of these two spectra remains a great similarity. This feature may be due to the facts that they have the same sequence and similar secondary structure because the CD spectrum was mainly influenced by the nearest neighbor and higher order sequence properties. Both of them show a strong positive band at 270 nm and a strong negative band near 242 nm. Rising the temperature, the band at 270 nm was reduced gradually. While the band at 242 nm keeps relative constant until 60 °C, an obvious decrease in the magnitude only occurs after that temperature, which was accompanied by a little bit blue-shift in the wavelength. At the temperatures above 80 °C, the CD spectra for the monomer and dimer overlap together roughly in the whole wavelength, and display a shape typical of a non base-pairing single-stranded RNA (shown as Figure 4.7 B) (Sosnick *et al.,* 2000). Given the spectra at different temperature, the positive band at 270 nm was probably related to the 5' region which formed the dimer, while the negative band at 242 nm more likely associate with the long stem-loop structure in the 3' end. Because after dimerization, the resulting intramolecular duplex in the 5' end is more stable than the corresponding structure in the monomer, hence more energy was needed to disrupt this region into a single strand. This analysis was well consistent with the CD spectral changes at 270 nm, because the magnitude for the monomer was decreased evenly from 20 to 80 °C, while for the dimer, this transition was relatively slow at the beginning and accelerated gradually.

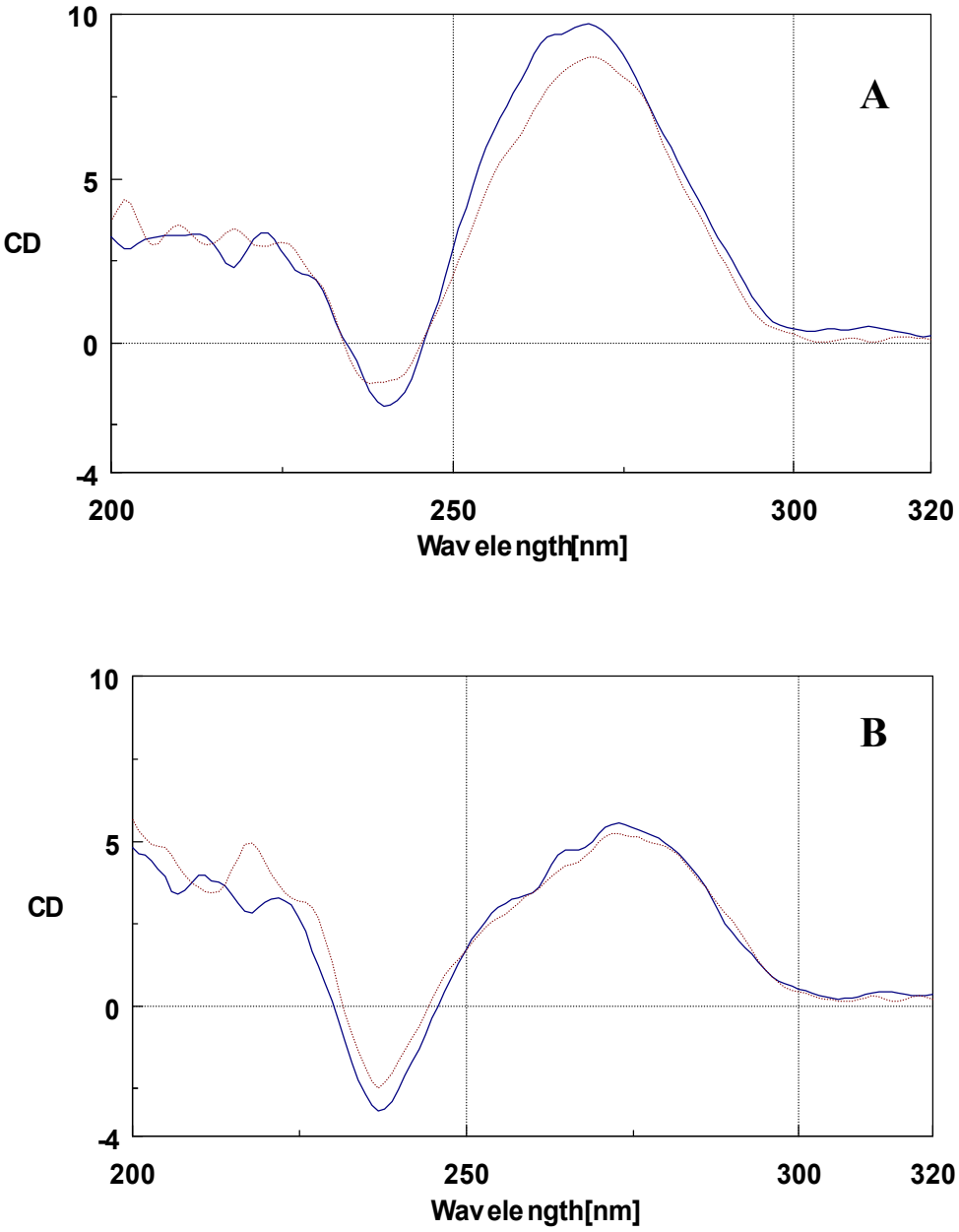


Figure 4.7. Circular dichroism spectra of Monomer (solid) and Dimer (dot) at 20 °C (**A**) and 80 °C (**B**); Both the monomer and dimmer contain 2 μM 79 mer RNA and in the buffer with 5 mM Na_2HPO4, pH 7.0, 100mM NaCl and 0.1mM EDTA. The monomer was made by heating at 85 for 3 mins and quick cool on ice for 30 min, and the dimer was made by adding $MgCl_2$ to the buffer, the final concentration is 10 mM.(it actually is a mixture of both monomer and dimmer according to the gel) The spectra were calibrated by their respective buffer.

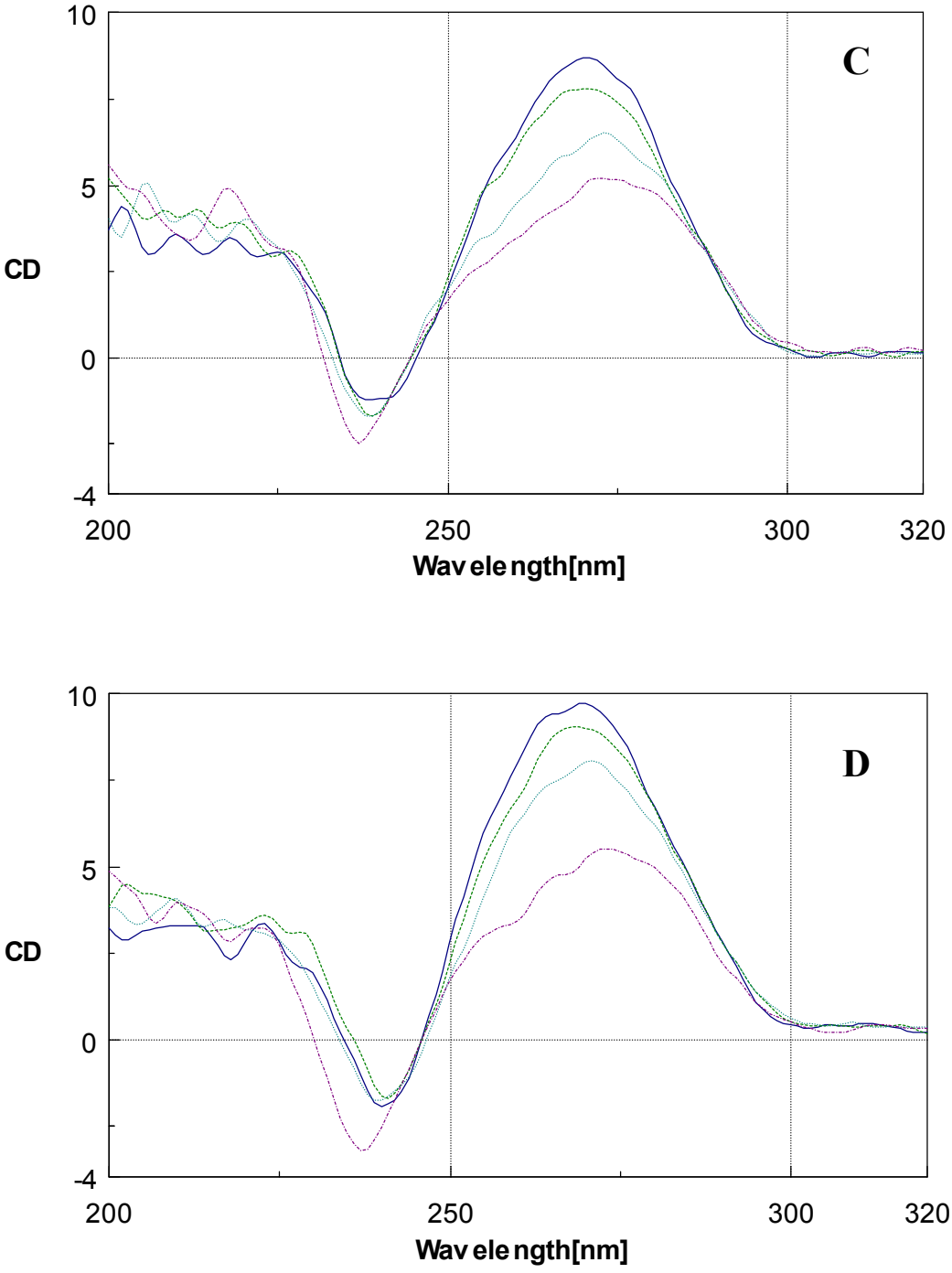


Figure 4.7 Circular Dichroism spectra of Monomer (**C**) and Dimer (**D**) as a function of temperature, namely 20 °C (solid) , 40 °C (dash), 60 °C (dot), 80 °C (dot-dash).

In contrast, the long stem-loop structure at 3' end contributes nothing to the dimeriztion. No structural rearrangement occurs during this process, so the tendency of the spectral change at 242 nm was quite similar between the monomer and dimmer. Moreover, this region is relatively stable than the 5' end, so it began collapsing into single strand until 60 °C.

RNase H assay

To probe more precisely the nucleotides involved in the single stranded and duplex segments of the dimmer conformation, a RNase H assay was carried out (Li and Wartell, 1998). Figure 4.8 shows the results of incubating 0.5 unit of RNase H at 25 °C for 30 min with 1.5 μM RNA and 9μM various DNA oligomer. The percentage of RNA degradation was shown in table 4.1. From that we can see the major difference between the digestion of monomer and dimer was obtained with the first two ODNs. Especially for ODN 1, the dimer shows a nearly two fold decrease in RNA degradation compared with the monomer. Given the predicted structures, this ODN was complementary to the single-stranded loop in the first stem-loop structure. The predicted dimer structure places the corresponding region in an intermolecular duplex. Thus, it can't be accessed by the ODN easily any more, which is consistent with reduced RNase H cleavage. This is also predicted for ODN2. Its complementary single stranded region in the monomer contains some paired bases in the predicted dimer structure. The reduction in the RNase H degradation for the dimmer is not so obvious as the former. But a 16 % difference is observed. Another fact worth of noting that although the target regions corresponding to ODN 1 and ODN 2 are both base paired, the percentage in the dimer degradation are not

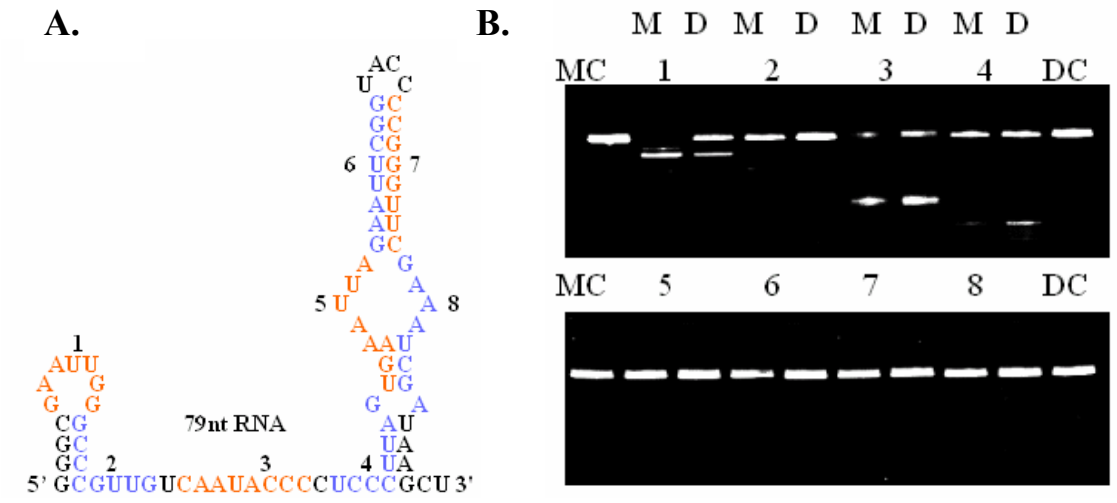


Figure 4.8 A 8 ODNs used for RNase H digestion, the red and blue color indicate the regions for these ODNs. **B** RNA degradation by RNase H. M: Monomer; D: Dimer; MC: Monomer Control; DC: Dimer Control, 1-8 correspond to the 8 ODNs used.

Table 4.1 RNA degradation by RNase H

ODNs	Sequence	% RNA degradation	
		Monomer	Dimer
1	5'-CCAATTCG-3'	70	34
2	5'-CAACGGGC-3'	38	22
3	5'- GGGTATTG-3'	55	60
4	5'-CTAAGGGA-3'	43	45
5	5'-TAATTTCA-3'	3	2
6	5'- CCGAATTC-3'	2	5
7	5'-GAACCCGG-3'	3	4
8	5'-TCGATTTC-3'	3	2

the same. One is 22%, the other is 34%. This was probably due to the number of the base pairs in the target regions. For the ODN 1 target region, there is a two- base bulge (AU) in the sequence, which was not present in the ODN 2 target region. Six ODNs targeted to the 3' portion of the RNA, no detectable difference was found between the monomer and dimer. All of them display a similar digestion pattern in both cases. This result was in good agreement with the predicted structural since the corresponding target region of these six ODNs was from position 21 to the 3' end which was predicted to have a similar secondary structure as a monomer or dimer. Therefore, the RNA can be degraded to a great extent in the presence of ODN 3 or 4. While when ODN 5-8 was added individually, it almost remains intact.

Hybridization Assay

Combining the results noted above, we found that the first and the second stem-loop region in the 79 nt RNA play a quite different role in forming the dimer. To some sense, the former served as an "actor", while the latter was just an "audience". It is known that the "actor" can put on a performance whether the "audience" is present or absent. If that is the case, the first stem-loop along with the partial single strand region should be capable of forming a dimer without the second stem-loop. This expectation was met. Two bands were observed by the 50 nt RNA which contains the first 50nt from the 5' end without a heating-cooling process. From structure predictions, it was found that this 50 nt RNA utilizes the same region as the 79 mer to form the dimer. An interaction occurs between these two molecules if they are just mixed together. This was demonstrated by the gel electrophoresis where an extra band was observed corresponding to a heterodimer

formed by the 50 nt and 79 nt RNA. In contrast, no evidence was found that the 47 nt RNA can bind to the 79 nt RNA and the 50 nt RNA. This means that the 47 nt lacks the essential part to form a heterodimer with these two molecules.

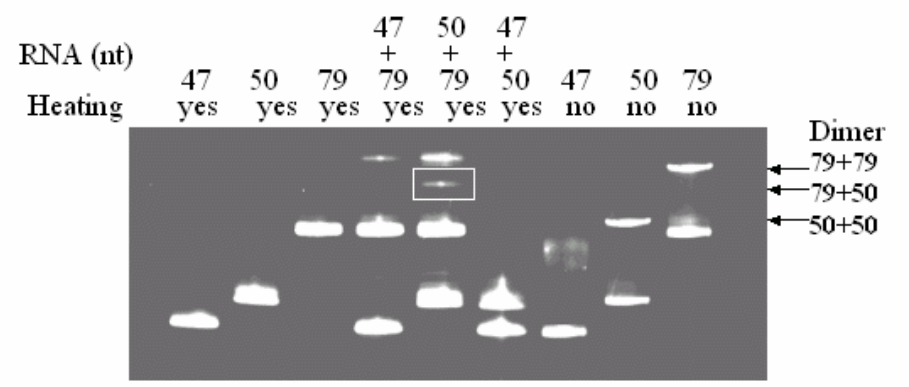


Figure 4.9 RNA hybridization of 79 nt RNA with 50 nt RNA and 47 nt RNA

Previous studies have shown that RNA conformation can be influenced by magnesium ion. In this study, we simultaneously examined the influence of both temperature and metal ion on multiple RNA conformations. The TGGE method was used to detect nucleic acid conformation transitions. The results show clearly that the RNA has multiple structures/conformations co-existing. The multiple RNA conformations were slowly changing to the stable high mobility monomer structure as the temperature increased. In the PAGE and TGGE analysis, when the RNA was quickly heated and cooled, only the fast moving group in the gel remained in both TBE and TBM buffer as comparing to the RNA without heated and cooled. Without quick heated and cooled, multiple bands in Group I and Group II will eventually became one band in fast-moving Group II locations at higher temperature in TGGE gel, indicating that there is mostly

single RNA molecule in the gel, while the RNA can fold into different structures in TBE and TBM buffer at low temperature.

The sedimentation equilibrium results support the interpretation that the RNA molecules form multimers at low temperature. The average molecule weight value of 1.72 times the monomer molecular weight in the TBE buffer and 1.6 in the 1xTBM buffer at 5 °C indicates the there is a significant population of dimer or higher multimers. The sedimentation equilibrium experiments at 28 °C showed that the RNA had a higher molecular weight (1.6x monomer) in TBM buffer than that in the TBE buffer (1.1x monomer) thus the RNA is mostly a monomer structure in TBE buffer at 28 °C, while in TBM buffer, the RNA contains a fraction of multimer structure. This is consistent with TGGE gel patterns in TBE and TBM buffer. This also indicates that the fast moving group II in the TGGE gel patterns in TBE and TBM buffer. This also indicates that the fast moving group II bands in the TGGE gel correspond to the RNA monomer structure, since only the fast moving band remains in the TGGE gel in TBE at 28 °C. Because the sedimentation equilibrium experiments also showed the RNA multimer formation at low temperature in both TBE and TBM buffer, this would indicate that the slow-moving bands in Group II in TGGE correspond to RNA multimer conformation in TBE and TBM buffers, In other words, the RNA monomer and multimer co-exist at low temperature.

The data indicates that divalent cations such as Mg^{2+} not only can influence RNA secondary structure but they can also influence the formation of RNA multimers. At low temperature conditions, we observed the divalent cations altered the ratio of RNA conformations in the TGGE gel by apparently stabilizing a few of them. This is likely due to interactions of the divalent cation with specific sites, however further work will be

needed to confirm this hypothesis. In the absence of divalent cations as in TBE buffer, there are multiple bands within group I or group II at low temperature. However, those multiple bands within the same group in the gel become one distinctive band in group I and group II in the presence of divalent cations. Since the bands in fast-moving group II correspond to the monomer while the bands in slow –moving group I correspond to RNA dimers, divalent cations can thus stabilized both a specific RNA monomer conformation as well as a dimer conformation. In TBE buffer, we observed a gradual decrease in the dimer bands as temperature increases in TGGE, while in TBM buffer, the ratio of dimer to monomer is quite stable at a wide range of low temperature from 4 °C to at least 28 °C. This is consistent with the results from the sedimentation experiments at 5 °C vs 28 °C in TBE and TBM buffers. A gradual decrease of the dimer bands in TBM buffer only started to appear at a much higher temperature at about 30 °C in TGGE gel, thus divalent cations also cause the RNA dimer to be less sensitive to temperature increase.

RNA structural prediction (mfold 3.1) suggests that the first 21 nt from the 5' end of the 79 mer are critical for dimerization. This region changes from an intramolecular stem-loop to an intermolecular duplex. A RNase H assay for accessible sites was carried out by using 8 ODNs. Only two ODNs show a different RNA degradation pattern between the monomer and dimer. The results were consistent with the single strand region predicted to be involved in dimerization. Hybridization studies involving segments of the 79 mer verified the essential role of this region for dimerization. CD spectra show differences around 270 nm between the monomer and dimer forms. The results were also consistent with the prediction. It may be that this provides an example of an initial loop-loop interaction that rearranges to maximize base-pairing.

References

Dripps, D., Wartell, R.M.,(1987) DNA bending induced by the catabolite activator protein allows ring formation of a 144 bp DNA.*J Biomol Struct Dyn*. **5**(1):1-13.

Green, M.R., (1986) Pre-mRNA splicing. *Annu Rev Genet*. **20**:671-708.

Inoue H, Hayase Y, Iwai S, Ohtsuka E. (1988) Sequence-specific cleavage of RNA using chimeric DNA splints and RNase H. *Nucleic Acids Symp Ser*. **19:**135-8.

Johnson KH, Gray DM. (1991) A method for estimating the nearest neighbor base-pair content of RNAs using CD and absorption spectroscopy. *Biopolymers*. **31**(4):373-84.

Joyce, G. F., (1996) Building the RNA world. Ribozymes. *Curr Biol*. **6**:965-7.

Koslowsky, D.J., (2004) A historical perspective on RNA editing: how the peculiar and bizarre became mainstream. *Methods Mol Biol***. 265**:161-97.

Lavorgna, G., Dahary, D., Lehner, B., Sorek, R., Sanderson, C. M., Casari, G. (2004) In search of antisense. *Trends Biochem Sci*. **29**:88-94.

Li J., Wartell R.M., (1998) RNase H1 can catalyze RNA/DNA hybrid formation andcleavage with stable hairpin or duplex DNA oligomers. *Biochemistry* **37**:5154-61.

Paillart JC, Marquet R, Skripkin E, Ehresmann C, Ehresmann B. (1996) Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie*. **78**:639-53.

Sosnick TR, Fang X, Shelton VM. (2000) Application of circular dichroism to study RNA folding transitions. *Methods Enzymol*. **317**:393-409.

Storz G, Opdyke JA, Zhang A. (2004) Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol*. **7**:140-4.

Zuker, M., Mathews, D.H., Turner, D.H. (1999) Algorithms and Thermodynamics for RNA secondary structure prediction: A practical guide in RNA biochemistry and biotechnology, 11-43, Kluwer Academic Publishers.

VITA

Xueguang Sun was born in Harbin Heilongjiang, P.R China. He received his B.S degree in Analytical Chemistry from University of Science and Technology Beijing, China in 1995. He moved into the field of biological science in graduate study at Institute of Biophysics, Chinese Academy of Sciences (CAS) where he was a group member working on the National Ninth Five-Year Project of CAS. He published six papers in *J. Biomol. Struct. Dyn,, Biochem.& Mol.Bio.Intl., Chemistry Letters* and *Inorganic Chem. Commun*. *et al.* respectively and was awarded the CAS President Scholarship due to his productive work. In 1999, he received his first Ph.D. degree in Biophysics from CAS. Shortly after that, he went to the United States to pursue his second Ph.D. in molecular biology in School of Biology, Georgia Tech. From 2000 to 2005, he has been working under the supervision of Dr. Roger Wartell. His research has resulted in one publication in *Nucleic Acid Research* and will likely lead to one or two more journal publications. He expects to receive his Ph.D. from Georgia Institute of Technology in Fall, 2005.